the second PAH domain of Sin3

structural biology & NMR methodology

hugovaningen

The second PAH domain of Sin3

Structural biology and NMR methodology

een wetenschappelijke proeve op het gebied van de
Natuurwetenschappen, Wiskunde en Informatica

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de Rector Magnificus prof. dr. C.W.P.M. Blom,
volgens het besluit van het College van Decanen
in het openbaar te verdedigen op 20 september 2006
des namiddags om 1.30 uur precies
door

Hugo van Ingen

geboren op 30 november 1976
te Weurt

Promotor
prof. dr. C.W. Hilbers
Co-promotor
dr. G.W. Vuister

Manuscriptcommissie
prof. dr. Sybren Wijmenga
prof. dr. Henk Stunnenberg
dr. Frans Mulder, Rijksuniversiteit Groningen

ISBN-10: 90-9020810-0
ISBN-13: 978-90-9020810-7

Printed by PrintPartners Ipskamp, Enschede

Color figures can be viewed on-line at http://www.nmr.chem.uu.nl/~hugo/.

The research described in this thesis was performed at the Department of Physical Chemistry/Biophysical Chemistry, Institute of Molecules and Materials, Radboud University Nijmegen, The Netherlands. This work was supported by grant JC-99-03 to G.W.V. from the Netherlands Organisation for Scientific Research, Division for the Chemical Sciences (NWO-CW).

I-I-I-I... this could be a-a-a-a lot more,
uh, uh, uh, uh, uh, uh, complex,
I mean, it's not just, it might not be just such a simple... uh, you know?

The Dude - The Big Lebowski

Well, I'm sure they'll be able to fix it. You can't stop modern science. Can't stop it, you can't stop it. Can't stop science. Can't be stopped, no way, no how, science just marches--

George Costanza - Seinfeld

voor mijn ouders

Table of contents

List of abbreviations

1D	one-dimensional
2D	two-dimensional
3D	three-dimensional
bHLHzip	basic region-helix-loop-helix-leucine zipper
CPMG	Carr-Purcell-Meiboom-Gill
COSY	correlation spectroscopy
DNA	deoxyribonucleic acid
FID	free induction decay
HAT	histone acetyl transferase
HCR	highly conserved region
HDAC	histone deacetylase
HID	HDAC interacting domain
ITC	isothermal calorimetry
kDa	kiloDalton
MD	molecular dynamics
MQ	multiple quantum
MRI	magnetic resonance imaging
mRNA	messenger RNA
NCoR	nuclear co-repressor
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect spectroscopy
NuRD	nucleosome remodeling and histone deacetylase complex
OGT	O-linked N-acetylglucosamine transferase
PAH	paired amphipathic helix, or poly amphipathic helix
PDB	protein data bank
PFG	pulsed field gradient
RbAp	retinoblastoma associated protein
RDC	residual dipolar coupling
RF	radio-frequency
RMSD	root-mean-square-deviation
RNA	ribonucleic acid
SAP	Sin3 associated protein
SDS3	suppressor of defective silencing 3
SE	sensitivity enhancement
SID	Sin3 interacting domain
Sin3	Swi5 independent protein 3
SMRT	silencing mediator for retinoic acid and thyroid hormone receptors
SPR	surface plasmon resonance
SQ	single quantum
TOCSY	total correlation spectroscopy

Chapter 1.

General introduction

Nuclear Magnetic Resonance (NMR) is a powerful technique to study the relationship between the structure, dynamics and function of proteins. This thesis presents an NMR study on the second PAH domain of a protein called Sin3, which is part of a special group of proteins that is ultimately responsible for shaping an organism by regulating the expression of its genetic information. In this chapter, a general overview of the biological role of the Sin3 protein and the role of its PAH domains is given (section 1.1). Furthermore, the role of NMR in structural biology is discussed (section 1.2) and basic NMR theory is introduced (section 1.3).

1.1 Sin3: regulator of gene transcription

Even the smallest unit of life, the cell, is staggeringly complex. The cell is composed of a multitude of molecules of widely varying types and sizes that act either by themselves or by associating into higher-order supramolecular assemblies. The cell's life cycle relies on the complex interplay between all these constituents and, importantly, on the interaction with its surroundings. Proteins are the 'workhorses' of the cell, carrying out most tasks like signal transduction, transport of metabolites over the cell membrane, catalysis of reactions, and the regulation of these and many other processes. They are continuously synthesized, transported, used and degraded.

Proteins are produced, or expressed, via a two-step mechanism: transcription and translation. In short, a gene, which codes the amino acid sequence of the protein, is first transcribed by a complex of several proteins, referred to as the transcription machinery, into messenger RNA (mRNA). Then, the mRNA is transported from the nucleus to the endoplasmatic reticulum in the cytoplasm, where it is translated into the corresponding protein by a complex of proteins and RNA molecules, called ribosomes (reviewed in Orhinades and Reinberg, 2002). The focus of this thesis is the protein Sin3, which is involved in the regulation of gene transcription. In the following subsections, I will first describe the general mechanism of the regulation of gene transcription in eukaryotes. Then, I will introduce the Sin3/HDAC complex and discuss its biological role. Finally, I focus on the Sin3 protein, its

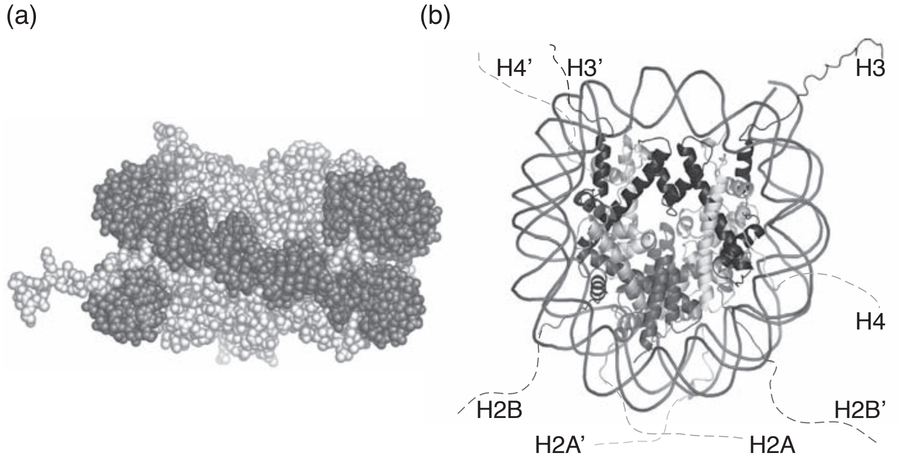


Figure 1.1. Structure of the nucleosome core particle (Luger et al., 1997; PDB code 1AOI). (a) Space filling representation of the nucleosome, viewed perpendicular to the symmetry axis of the superhelical DNA. The DNA, shown in dark gray, is wrapped around the histone octamer, shown in light grey. (b) Ribbon representation of the nucleosome core particle, viewed from the top along the superhelical symmetry axis. The N-terminal tails of the four histone proteins H2A, H2B, H3 and H4 are schematically indicated by the dashed lines. The two copies of each histone pair are distinguished as unprimed and primed copies, where the histone fold of the primed copy is primarily associated with the bottom half of the DNA and the unprimed copy with the top half.

PAH domains and the structural aspects of the interaction between the PAH domains and other proteins.

General mechanism of gene transcription regulation

Tight regulation of the expression of proteins and their distribution, both spatially and temporally, is essential for proper functioning of all cells. To this end, several regulatory mechanisms are used that target different stages of protein expression. Roughly, these regulatory processes can affect gene transcription, mRNA translation or directly regulate protein activity by post-translational modifications. Regulation of the initiation of gene transcription is one of the key processes and is central for all organisms to both their development and homeostasis (Roeder, 2003). In eukaryotic organisms, gene transcription is typically up- or down-regulated via certain factors that bind to specific DNA sequences upstream of the gene to be regulated. The set of targeted genes is determined by the cell type, the status of the cell in its development, its stage in the cell-cycle and by the demands raised by the environment of the cell. The regulatory factors then recruit co-regulatory factors that either directly interact with components of the transcription machinery or alter the accessibility of the DNA to the transcription machinery (reviewed in Roeder, 2005).

Eukaryotic DNA is not naked as in bacteria, but is wrapped around proteins and highly compacted into a structure called chromatin (reviewed in Felsenfeld and Groudline, 2003). It has long been known that highly condensed, inaccessible chromatin, ‘heterochromatin’, is usually not transcribed, while less condensed regions, ‘euchromatin’, correlate with actively transcribed regions of the chromosome. Recently, it has also become clear that activation and repression of gene transcription is closely related to reversible modifications of the smallest structural unit of chromatin, the nucleosome (Fischle et al., 2003). It comprises a protein-DNA complex of circa 150 base pairs DNA wrapped around a protein core of two copies of each of the four histone proteins (denoted as H2A, H2B, H3 and H4) in a 1¾ superhelical turn (Figure 1.1; reviewed in Kornberg and Lorch, 1999). The N-terminal tails of the histone proteins protrude from the nucleosome surface and are accessible to enzymes that chemically modify specific amino acids that are highly conserved between different species. Typical modifications involve lysine residues that may be acetylated, methylated, or coupled to ubiquitin; arginine residues that may be methylated and serine residues that are phosphorylated. Together, these modifications constitute a complex code, called the 'histone code', that influences gene transcription (Juwelein and Allis, 2001; Berger, 2002).

Acetylation of histone tails has long been correlated with increased transcription (Allfrey et al., 1964). Rather than changing the nucleosome structure, acetylation appears to weaken the interaction between nucleosomes, enhancing the accessibility to transcription factors. Conversely, deacetylation of the histone tails results in increased condensation of the chromatin structure, reducing the accessibility and thus resulting in repression of gene transcription (Grunstein, 1997; Eberharter and Becker, 2002). Alternatively, acetylated histones may recruit chromatin remodeling proteins that enhance gene transcription by remodeling the nucleosome to expose the otherwise inaccessible promoter DNA (Logie et al., 1999; Mellor, 2005). Histones are acetylated by histone acetyl transferases (HAT) enzymes that use acetyl-coenzymeA to transfer an acetyl group onto a lysine residue, resulting in the neutralization of its positive charge (reviewed in Berger, 1999). The acetyl groups can be removed by histone deacetylases (HDAC) enzymes, which are grouped into three classes with distinct function based on their homology (reviewed in Yang and Seto, 2003). These enzymes are typically part of a multi-subunit complex called a co-activator (in the case of HATs) or a co-repressor (in the case of HDACs). These complexes are targeted to specific genes via their interaction with activator and repressor proteins, which specifically bind to distal regulatory elements upstream of the promoter of the target gene.

The cell uses a limited set of co-activator and co-repressor complexes to regulate the expression of its genes (reviewed in Featherstone, 2002 and Burke and Baniahmad, 2000). Some well-characterized examples of co-activators include the Mediator complex, that both mediates interaction between activator proteins and the general transcription factors in the transcription machinery (reviewed in Kornberg, 2005) and contains HAT activity (Lorch et al., 2000), the HAT proteins p300 and CBP (reviewed in Kalkhoven, 2004) that function as global regulators of activation and the chromatin remodeling complex Swi-Snf (reviewed in Sudarsunam and Winston, 2000). An example of a co-repressor complex is the nucleosome remodeling and histone deacetylase complex NuRD (also known as Mi-2) that combines both chromatin remodeling and deacetylation of histones to repress transcription (reviewed in Feng and Zhang, 2003). Another co-repressor is the nuclear receptor co-repressor NCoR that together with SMRT, the silencing mediator for retinoic acid and thyroid hormone receptors, represses transcription of genes targeted by the nuclear receptor family proteins (reviewed in Hu and Lazar, 2000). Here, I focus on the Sin3/HDAC co-repressor complex. which is recruited by many different repressors and is involved in the silencing of genes connected to regulation of the cell-cycle, embryonic development and the switch between cell proliferation and differentiation. I will first discuss the composition of this multi-subunit complex and then focus on its biological role.

The Sin3/HDAC co-repressor complex

The Sin3 protein was first discovered in yeast as a negative regulator of gene expression (Sternberg et al., 1987). The name Sin3 is derived from its discovery in a screen for Swi5 independent transcription factors. Later, it has been identified in many eukaryotic organisms. The Sin3/HDAC complex was initially discovered in 1997, reported in 7 independent research papers (Alland et al., 1997; Hassig et al., 1997; Heinzel et al., 1997; Kadosh and Struhl, 1997; Laherty et al., 1997; Nagy et al., 1997; Zhang et al., 1997). Additional studies have resulted in the identification of the core components of the co-repressor complex (see also Figure 1.2b).

Sin3 binds both the DNA binding repressor proteins and the Class I histone deacetylating proteins, HDAC1 and HDAC2. Thereby, it establishes the crucial link between the target gene and the chromatin modifying activity. It is also the docking platform for several other proteins, including the Sin3 associated proteins (SAP18 and SAP30), which are thought to stabilize the complex (Zhang et al., 1997; Zhang et al., 1998) and the histone targeting proteins (retinoblastoma associated protein RbAp48 and RbAp46), which stabilize the contact with the nucleosome (Hassig et al., 1997; Zhang et al., 1997). Additionally, the core Sin3/HDAC co-repressor complex consists of SDS3 (suppressor of defective silencing 3, Alland et al., 2002), SAP130 and SAP180 (Fleischer et al., 2003). The core co-repressor complex can be extended with extra catalytic modules, such as the monosaccharide O-Glc-Nac transferase OGT (Yang et al., 2002), which transfers a O-linked N-acetylglucosamine to serine or threonine residues of other proteins to regulate their activity. It is proposed that O-GlcN-acetylation of RNA polymerase II results in inhibition of transcription elongation. Thus, the Sin3 complex might function to regulate both the initiation and the elongation of transcription. Also, Sin3 interacts with a methyltransferase that methylates specific lysine residues in the histone tails, thereby promoting condensation of the chromatin into a heterochromatin state (Yang et al., 2003). Sin3 has also been found to associate with the methyltransferase MeCP2, which methylates CpG dinucleotides (Jones et al., 1998; Nan et al., 1998). Methylation of DNA is an epigenetic modification widely used in the vertebrate genome and plays an important role in imprinting, which is the monoallelic silencing of a gene. It also is important for the inactivation of the X-chromosome and embryonic development.

The biological role of Sin3/HDAC repression

Sin3/HDAC mediated repression is essential for cellular survival. Knock-out studies in mice have shown that deletion of Sin3 causes embryonic death (Dannenberg et al., 2005; Cowley et al., 2005). Conditional deletion of Sin3 in mammalian cell cultures showed a significant up-regulation of transcription of up to 900 genes (Dannenberg et al., 2005). These knock-out studies showed that Sin3 is essential in embryogenesis, cell cycle regulation and is required for the development and homeostasis of cells in the lymphoid lineage (Dannenberg et al., 2005; Cowley et al., 2005). Via its association with DNA methyltransferases and SDS3, which has been shown to be required for pericentric heterochromatin formation, Sin3 plays a crucial role in maintaining genomic stability (David et al., 2003; Silverstein et al., 2003; Xin et al., 2004).

Examination of Sin3/HDAC recruiting repressors and their target genes results in a more detailed understanding of the biological role of Sin3/HDAC mediated repression. The interaction between Sin3 and the transcriptional repressor Mad1 has been particularly well studied (Ayer et al., 1995; Schreiber-Agus et al., 1995). The Mad family proteins (Mad1, Mad2/Mxi1, Mad3, Mad4 and Mnt) are part of the Myc/Max/Mad network that is involved in the switch between cell proliferation and cell differentiation in mammals (Adhikary and Eilers, 2005; Grandori et al., 2000). The proto-oncogene product Myc and the proteins Max and Mad are basic region-helix-loop-helix-leucine zipper (bHLHZip) transcription factors that bind as heterodimers of either Myc/Max or Mad/Max to the hexanucleotide CACGTG, the so called E-box element. Binding of Myc/Max results in the recruitment of histone acetyl transferase, transcriptional activation of target genes and stimulation of cell proliferation, while Mad/Max recruits the Sin3/HDAC co-repressor complex, resulting in gene silencing and the start of cell differentiation. Deregulation of Myc/Max induced transcriptional activation has been linked to the development of many human cancers.

Loss of Sin3/HDAC activity has also been implicated in the development of breast malignancies via the repressor and tumor suppressor SMAR1 that targets the cyclin D1 gene, a cell cycle regulator (Rampalli et al., 2005). The role of Sin3 in tumorigenesis is further emphasized by its association with various tumor suppressor proteins such as p53 (Murphy et al., 1999; Tanikawa et al., 2004), p33ING1 (Kuzmichev et al., 2002) and Menin (Kim et al., 2003). Furthermore, the Sin3/HDAC complex also plays an important role in regulation of the cell cycle as it is recruited to genes essential for DNA synthesis and cell cycle progression via the E2F family of transcription factors (Lai et al., 2001; Rayman et al., 2002). Finally, aberrant Sin3/HDAC mediated repression has been linked to several forms of leukemia (David et al. 1998; Lin et al., 1998; Kahn et al., 2001), in which fusion proteins resulting from chromosomal translocations interact too strongly with HDAC complexes, keeping target genes in a repressed state and thus blocking normal blood cell development.

Sin3: domain organization

Sin3, the central protein in the Sin3/HDAC co-repressor complex, is a large protein that is highly conserved in eukaryotic organisms. Sin3 proteins have ~1000-1500 residues and have been identified in various fungi, plants and animals, including mammals. Some organisms have multiple forms of the Sin3

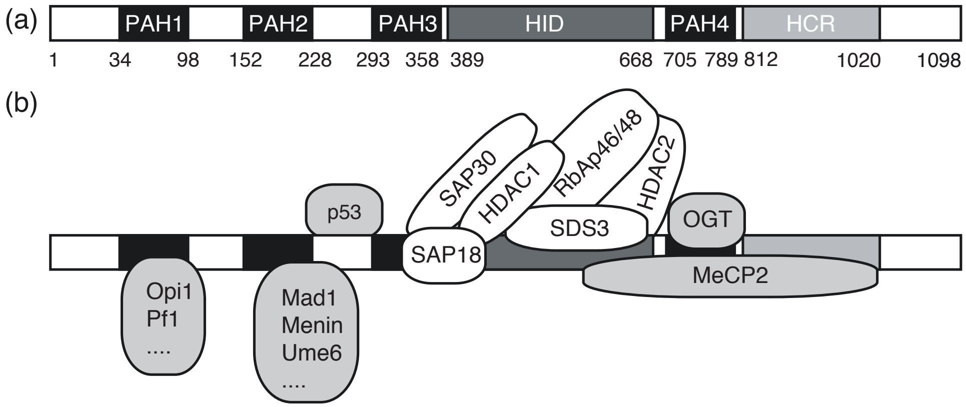


Figure 1.2. (a) domain organization of Sin3 (PAH = poly amphipathic helix domain; HID = histone deacetylase interacting domain; HCR = highly conserved region). Residue numbering corresponds to the mouse mSin3B isoform (SwissProt accession code Q62141). (b) Schematic representation of the Sin3/HDAC complex. Components of the core complex are shown in white, additional components of the complex and the Sin3 recruiting repressors are shown in grey. Core components interact with Sin3 around PAH3 and the HID domain, while the additional factors interact with Sin3 in several different regions (adapted from Silverstein and Ekwall, 2005).

protein, such as fission yeast, which contains the three isoforms Pst1, Pst2 and Pst3 (Dang et al., 2002, Silverstein and Ekwall, 2005). Mice and humans have two isoforms, termed mSin3A and mSin3B (Halleck et al., 1995; Ayer et al., 1995). Isoform mSin3B has a shorter N-terminal region than mSin3A, but has an otherwise highly similar sequence. In spite of this similarity in sequence, recent studies have suggested that mSin3A and mSin3B most likely have separable functional roles (Dannenberg et al., 2005; Cowley et al., 2005), however, the nature and extent of these differences are currently not known. To add further to the complexity, several organisms contain different Sin3 splice variants, whose functional role remains to be established (Alland et al., 1997).

Sin3 contains multiple, conserved protein-protein interaction domains to bind its many partners (Figure 1.2a). Originally, four imperfect repeats, denoted as the PAH domains (PAH1-4), were identified that were presumed to fold into a pair of amphipathic helices (Wang et al., 1990). In addition, a highly conserved region between the third and fourth PAH domain, referred to as the HDAC interaction domain (HID), is responsible for interaction with HDAC1/2. The C-terminal end of Sin3 contains a highly conserved region, the HCR domain. The function of the HCR domain is still unclear but it has been shown necessary for Sin3 repressor function (Laherty et al., 1997).

The PAH domains are typically identified as the regions required for interaction with the repressor. In their recent review of Sin3, Silverstein and Ekwall (2005) summarize the required interaction surfaces of all identified Sin3 interacting proteins. Most repressors require the second PAH domain, PAH2, and to a lesser extent the PAH1 and PAH3 domains as interaction surfaces. The PAH4 domain is only required for association with the monosaccharide transferase OGT (Yang et al., 2002). For a number of interacting proteins, regions other than the PAH, HID or HCR domains were essential for binding as in the case of p53 (Murphy et al., 1999) and MeCP2 (Jones et al., 1998; Nan et al., 1998). New Sin3 interacting proteins are being identified on a regular basis, indicating the extensive role of Sin3 in transcriptional regulation.

Structural biology of Sin3-repressor interaction

Structurally, Sin3 has remained largely uncharacterized. Both the PAH domains and the HID and HCR domain share little homology with other protein domains. Precise identification of the interaction domains of Mad1 and Sin3 prompted structural characterization of this interaction (Ayer et al., 1995; Schreiber-Agus et al., 1995). The minimal Sin3 interacting domain (SID) of Mad1, necessary and sufficient for interaction with Sin3, is formed by an amphipathic helix of 13 amino acids in its N-terminal region, NIQMLLEAADYLE. This sequence binds with high affinity and specificity to the PAH2 domain of mSin3A and mSin3B (Eilers et al., 1999). The hydrophobic side of this amphipathic helix is the interaction surface for the PAH2 domain. As identification of the SID in most other Sin3 recruiting repressors has been not as clear-cut as for Mad1, the interaction between the Mad1-SID and the PAH2 domain has become the paradigm of Sin3-repressor interaction. The first detailed structural characterization of Sin3 was obtained from the solution structures of the PAH2 domain complexed to the SID of Mad1 which showed the complex to be folded as a ‘wedged helical bundle’, where the α-helix of the SID is inserted at angle of ~45° into the four helix bundle conformation of the PAH2 domain (Spronk et al., 2000; Brubaker et al., 2000; Figure 1.3). The PAH2-SID interaction is predominantly hydrophobic, with bulky and small residues in the SID complementing those in PAH2 as ‘knobs-into-holes’ (Spronk et al., 2000;

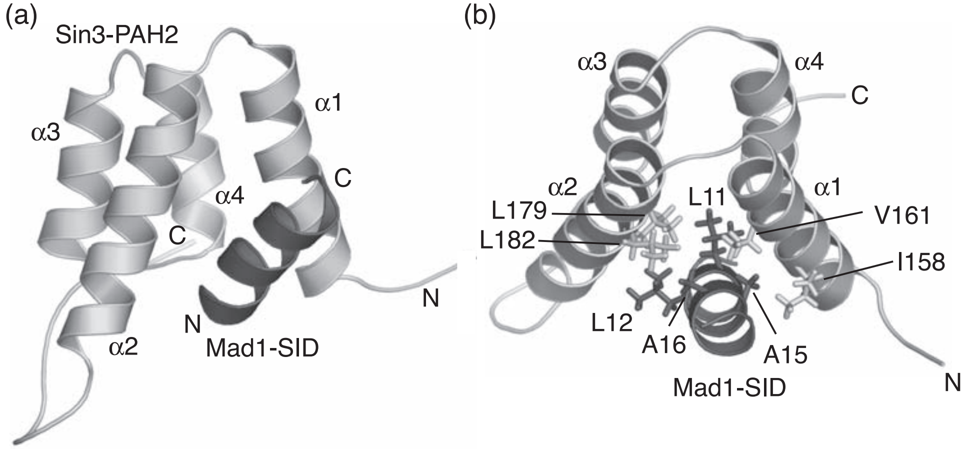


Figure 1.3. Solution structure of the Mad1-Sin3 interaction: the complex between a minimal 13-residue SID of Mad1 and the PAH2 domain of Sin3 (Spronk et al., 2000; PDB code 1E91). (a) Cartoon representation showing that the four helix bundle of the PAH2 domain and the amphipathic helix of the SID form a complex with a wedged-helical-bundle topology. The SID is shown in dark grey, the PAH2 domain in light grey. The four helices of PAH2 are labeled α1, α2, α3 and α4. (b) The interaction between the Mad1-SID and the PAH2 domain is mainly hydrophobic where bulky and small residues in the interface complement each other, like 'knobs-in-holes'. The bulky hydrophobic residues I158 and V161 in the α1 helix and L179 and L182 in the α2 helix are complemented by residues L11, L12, A15 and A16 of Mad1.

Brubaker et al., 2000; Cowley et al., 2004; Figure 1.3b).

These solution structures provide a structural basis for determining the principles of protein-protein interaction and selectivity involving the PAH domains. Two important questions, however, remain. The first question concerns the specificity of the PAH2 domain towards its target proteins. Although some PAH2 interacting proteins have similar tentative SIDs, many others lack a clear similarity with the Mad1-SID. Are these dissimilar SIDs bound in the same manner as Mad1 resulting in a lower binding affinity, or is the PAH2 domain highly adaptive such that it can bind unrelated sequences by forming structurally distinct interaction surfaces? The second question concerns the specificity of the PAH2 interacting protein. Sequence alignment of the four PAH domains shows that they have a high degree of sequence similarity suggesting that they have similar folds. Sequence similarity is not simply confined to residues that constitute the hydrophobic core, but also includes residues in the putative binding pockets. What then are the factors determining the selectivity towards a single PAH domain?

In this thesis, I focus on the adaptivity of the PAH2 domain of Sin3 by studying this domain in its bound and unbound state (discussed in Chapter 5). Comparison of these two states will indicate whether the PAH2 domain adjusts its fold to the SID or is able to present multiple, differently shaped binding pockets to the SID, or simply lacks any adaptive behavior. Structural comparison of these two states will allow a clear definition of the conformational changes occurring upon complex formation. Analysis of protein dynamics will discriminate between the induced fit (Koshland, 1958) and conformational selection mechanisms as the cause for these structural changes. In the case of a rigid, well-defined structure of the unbound state, it is most likely that the changes are induced as a consequence of, and thus after, initial complex formation. On the other hand, conformational heterogeneity in the unbound state would point to a mechanism where the complex is only formed with those conformations that match the interaction surface of the binding partner (conformational selection, based on the 'lock-and-key' concept of Fischer, 1894).

Furthermore, I elaborate on the Mad1-Sin3 interaction by studying the interaction between an extended Mad1-SID and the PAH2 domain (discussed in Chapter 4). An extended SID of 24 residues was observed to bind roughly 10-fold stronger to the PAH2 domain compared to the minimal thirteen-residue SID. This suggests that residues in the extension also contribute to the interaction surface. Identification of all interaction residues is not only important for an accurate description of the interaction between the Mad family proteins and the PAH2 domain, it will also help to explain the selectivity of this interaction.

1.2 NMR & Structural Biology

Fundamental understanding of protein function requires detailed knowledge of the interactions of the protein with its surroundings, i.e. a substrate, other proteins or even the solvent. These interactions critically depend on the three-dimensional (3D) structure of the protein, which defines the shape of the interaction surface and its chemical properties, i.e. its hydrophobicity or electrostatic properties. Furthermore, since structures are not static, interactions can cause structural rearrangements or sometimes require appropriate changes in the structure. The exploration of the relationship between structure, flexibility and function is central to structural biology. Research in this field on proteins is aimed at determining structures at atomic detail and the characterization of protein dynamics at all relevant time scales. This knowledge not only advances our fundamental understanding of Nature, it also provides opportunities to improve the treatment of human disease. In an approach called structure based drug discovery, knowledge of the protein structure is used to guide the design or improvement of pharmaceuticals (reviewed in Congreve et al., 2005). For instance, the aberrant repression of genes in leukemia can be relieved by small molecules that inhibit the HDAC enzymes (Melnick and Licht, 2002). Knowledge of the 3D structure of the HDAC enzyme will be critical for the development of highly specific inhibitors, which will likely have better pharmaceutical properties.

The relationship between the protein's amino acid sequence and its 3D shape, or 'fold', is still unknown. This is known as the protein-folding problem. However, it is known that highly similar sequences will have highly similar folds (Rost, 1999), and this is the basis of structure prediction by homology modeling. This approach is only successful if the fold is known, for instance it is contained in the current-day database of structures, the Protein Data Bank (PDB) (Berman et al., 2003). For most proteins experimental techniques are needed to solve the 3D structure. The two main experimental structure determination techniques used are X-ray crystallography and <u>n</u>uclear <u>m</u>agnetic <u>r</u>esonance (NMR). X-ray crystallography requires crystallization of the protein of interest and once properly diffracting crystals are obtained, structure determination is usually fast and typically results in very high quality structures. Proteins of high molecular mass are routinely solved by X-ray crystallography and even higher order complexes like the nucleosome are amenable (Luger et al., 1997). However, crystallization can be difficult or sometimes impossible to achieve.

Although NMR can nowadays also be used to solve protein structures in the solid state, typically a solution containing the protein of interest is used. Optimization of solvent conditions is usually faster than crystallization; however, data acquisition and analysis is not as fast for NMR as for X-ray. Much effort is put in the automation of data analysis to achieve a significant time gain (reviewed in Altieri and Byrd, 2004). NMR is particularly well suited for structure determination of proteins and complexes of limited molecular mass (< 40 kDa). In a recent review of the use of X-ray and NMR in structural genomics projects it was found that these techniques are complementary (Snyder et al., 2005; Yee et al., 2005). Proteins that do not form crystals are usually amenable in the solution state, while proteins that are not stable in solution do tend to give crystals.

Beyond structures: dynamics and interactions

Uniquely, NMR can be used not only to determine the atomic, three-dimensional structure of proteins, it can also provide a detailed characterization of protein dynamics at multiple time scales (for review see Palmer, 1997; Kay; 1998; see Figure 1.4). Dynamic processes in proteins cover a large range of time scales, from the very fast fluctuations of individual atoms on a pico second time scale, to correlated motions of secondary structure elements on micro- to millisecond time scale, to breathing motions of the global structure on a timescale slower than seconds (for review see Kern and Zuiderweg, 2003; Gurd and Rothgeb, 1979; Frauenfelder and Gratton, 1986). These motions are thought to be important for protein function in processes such as enzyme catalysis, ligand binding, protein folding and signal transduction.

Additionally, NMR is well suited for the study of interactions between biomolecules for a wide range of binding constants. An identification of the interaction surface is relatively easily obtained. Full structure determination of complexes is also possible, providing a wealth of information about the details of molecular recognition. Comparison of the structures of complexes with the structures of the free states of the interacting molecules, gives insight into the binding process, e.g. the importance of the induced fit mechanism. Furthermore, analysis of the dynamics of the molecules in the free and bound state can give insight in the functional significance of protein motions.

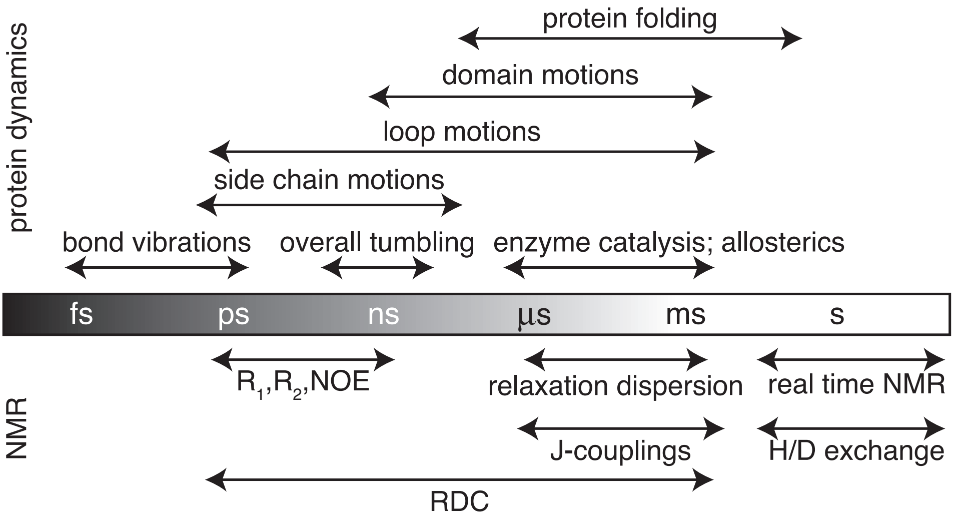


Figure 1.4. Overview of protein dynamics on different time scales and NMR parameters with their corresponding time scales.

In the studies reported in this thesis, I used NMR to study the PAH2 domain and its complex with the Mad1-SID. These protein domains are relatively small (12 and 3 kDa, respectively) and thus well suited for NMR. Furthermore, the PAH2 domain contains highly flexible parts that will likely complicate crystallization. Most importantly however, since this study is focused on both the unbound and bound state of the PAH2 domain and on the significance of dynamics in protein-protein interaction, the unique properties of NMR can be fully exploited.

This research not only includes the application of NMR in the context of structural biology but is also focused on the development and improvement of NMR methodology. On the one hand this is sometimes needed to solve a problem in case routine experiments fail. On the other hand, new methods can provide new information and thus result in a better characterization of the studied system. In the last section of this introduction, the basic principles of NMR are described, with special attention to the areas to which my work contributed.

1.3 NMR theory[†]

Since the first successful observation of nuclear magnetic resonance by Bloch and co-workers and, independently, Purcell and co-workers (Bloch et al., 1946; Purcell et al., 1946), NMR has become one of the most powerful techniques to study the chemical structure of matter. It relies on the magnetic moments of atomic nuclei, resulting from a quantum mechanical property of the atomic nucleus, called spin. Not all atoms have such a nuclear magnetic moment, but the most abundant nucleus in biomolecules and organic molecules, the hydrogen nucleus, possesses a spin. The abundance of the hydrogen spin in the human body has allowed the application of NMR in the health sciences in the form of nuclear magnetic resonance imaging (MRI).

When placed in an external magnetic field, the magnetic moments associated with the spin become partially aligned with the external field, creating a net overall magnetic polarization. For nuclear spins with spin quantum number ½, this net polarization can be thought to result from a differential population of the two allowed spin quantum states, α and β. In thermal equilibrium, the relative population of these states is determined by the energy difference between the two states according to the Boltzmann-equation. This energy difference is quite small which explains the inherent low sensitivity of NMR; only few nuclear spins (ca. 30-60) out of a million contribute to the net polarization at commonly used magnetic field strengths for biomolecular NMR. This small energy difference also means that the resonant electromagnetic radiation that is used to create an observable signal, has a low energy content and can be considered non-invasive. The frequency of this radiation is in the same range as radio waves (MHz) and can safely be used in *in vivo* applications such as MRI.

Quantum mechanical description of NMR

Rational design of new NMR experiments requires knowledge of how the nuclear spin states are transformed by the NMR experiment. The average nuclear spin state over the complete ensemble of spins in the sample is described by the spin density operator $\hat{\rho}$. Its time evolution is given by Liouville-von Neumann equation, which is derived from the time dependent Schrödinger equation:

$$i\hbar \frac{d}{dt}\hat{\rho}(t) = \hat{\hat{H}}(t)\hat{\rho}(t) \qquad [1.1]$$

[†] A number of textbooks describe the theory of NMR in depth, including the seminal works by Abragram (Abragram, 1961) and Ernst (Ernst et al., 1987) and works more suitable for the novice in the field by van de Ven, Cavanagh and Levitt (van de Ven, 1994; Cavanagh et al., 1996; Levitt, 2001).

in which $\hbar$ represents $h/2\pi$, where h is Planck's constant and $\hat{\hat{H}}$ is the Hamiltonian superoperator, which represents all physical interactions of the nuclear spins. In general, the Hamiltonian for a spin $\vec{I}_j$ with spin quantum number ½ contains five terms:

$$\begin{aligned}\hat{H}_j(t) = & -\gamma_j \vec{I}_j \vec{B}_0 - \vec{I}_j \bar{\bar{\delta}}_j(t) \vec{B}_0 \\ & + \sum_k \vec{I}_j \bar{\bar{D}}_{jk}(t) \vec{I}_k \\ & + \sum_k \vec{I}_j \bar{\bar{J}}_{jk}(t) \vec{I}_k - \gamma_j \vec{I}_j \vec{B}_1(t)\end{aligned} \qquad [1.2]$$

The first term represents the Zeeman interaction between the nuclear spin $\vec{I}_j$ and the external static magnetic field B_0 which also depends on the gyromagnetic ratio of the spin, γ_j and defines the characteristic resonance frequency of the nuclear spin, the Larmor frequency. The second term is the chemical shift interaction, where $\bar{\bar{\delta}}_j(t)$ is the chemical shift tensor. This interaction is caused by the small magnetic field induced by the electrons surrounding the nucleus. The direct dipole-dipole interaction is the third term in the Hamiltonian, where $\bar{\bar{D}}_{jk}(t)$ is the tensor describing the dipole-dipole interaction between spin $\vec{I}_j$ and $\vec{I}_k$. This dipolar coupling results from the magnetic interaction between the magnetic dipole moments of two spins. The fourth term is the scalar coupling or *J*-coupling interaction, where $\bar{\bar{J}}_{jk}(t)$ is the scalar coupling tensor between spin $\vec{I}_j$ and $\vec{I}_k$. Scalar coupling arises from the indirect magnetic interaction between two nuclear spins via the electrons in the connecting chemical bonds. The final term represents the interaction of the spin with the external radio frequency radiation that is used to perturb the system. Here, $\vec{B}_1$ is the resonant magnetic field component of the radio-frequency (RF) irradiation. The chemical shift, dipolar coupling and scalar coupling represent the interaction of the spin with its direct environment and are exploited in NMR experiments to derive structural and dynamical information about the molecules under study.

Chemical shift, J-coupling and dipolar coupling

When considering NMR applied to molecules in isotropic solution, the Hamiltonian can be separated in a time independent and a time dependent part. The time independent part is sufficient for the analysis of NMR experiments and can be calculated from the time average or, equivalently, the motional average of the full Hamiltonian. For isotropic liquids, all molecular orientations are equally likely, resulting in the isotropic average for the Hamiltonian: the chemical shift interaction is averaged to the isotropic chemical shift, the scalar coupling is averaged to the isotropic *J*-coupling constant, while the dipolar coupling is averaged to zero.[‡] The isotropic chemical shift slightly changes the Larmor frequency and depends very critically on the chemical environment of the nucleus, resulting in unique resonance frequencies for almost every spin in a molecule. Importantly, this differential shift forms the basis of the success of NMR as a tool for structure elucidation (reviewed in the context of biomolecular NMR in Case, 1998).

The isotropic *J*-coupling constant is sensitive to the local molecular conformation (Karplus, 1963) and results in a splitting of the resonance line in a multiplet structure. This splitting can be removed by decoupling techniques in which the B_1 magnetic field is used to either cancel the evolution of the density operator due to *J*-coupling or to suppress this evolution altogether. Decoupling of heteronuclear *J*-coupling during acquisition of the NMR signal is usually achieved via cyclical phase modulation of RF irra-

---

[‡] Long range intermolecular dipolar couplings between molecules separated by more than a diffusion sphere are not averaged to zero. These couplings, however, are very small and can safely be neglected here.

diation (Levitt and Freeman, 1981), which, when improperly implemented, can result in unexpected artifacts (discussed in Chapter 2). Furthermore, the *J*-coupling interaction is typically exploited in solution state NMR as a mechanism to transfer magnetization from one spin to another. Ultimately, this allows the assignment of all resonances to a particular nuclear spin of the molecule, which is the first step in the process of structure determination.

The time dependent part of the Hamiltonian causes rapid, randomly fluctuating local magnetic fields which can induce transitions of the nuclear spin state. After a perturbation of the nuclear spins, these transitions reestablish thermal equilibrium in a process called relaxation. The characteristic time scale of this relaxation process is determined by the global molecular rotational and translational diffusion and the local internal molecular motions. Thus, NMR relaxation parameters are very sensitive indicators of molecular dynamics. The dipolar interaction and the chemical shift interaction are the main relaxation mechanisms for most nuclei in biomolecular solution NMR. The influence of the dipolar interaction in relaxation is of critical importance for structure determination of biomolecules, as it allows transfer of magnetization from one to spin to a neighboring spin. This is the basis of nuclear Overhauser effect spectroscopy, NOESY, from which distances between nuclear spins can be derived (Jeener et al., 1979; Kumar et al., 1980). These distances can only be derived for spins that are relatively close in space (up to 5 Å). They are the main source of experimental restraints used in structure calculation of biomolecules. In addition, dipolar coupling can give rise to long-range structural information. When using an anisotropic liquid, the averaging of the dipolar couplings is not complete, resulting in residual dipolar couplings, RDCs, from which the angle between the internuclear vector and a fixed molecular frame can be calculated (reviewed in Bax and Grishaev, 2005). This information is very useful for structure determination and complements the short range distance information based on NOESY experiments.

Structure determination by NMR

Structure determination by NMR involves three basic steps, after preparation of a suitable sample: first, the assignment of every resonance to a nuclear spin in the molecule, second, the measurement of structural restraints and third, the calculation and subsequent refinement of a set structures that agree with the experimental restraints. For small proteins, assignments can be obtained using the sequential assignment approach (Wüthrich, 1986) based on homonuclear two-dimensional (2D) ^{1}H-^{1}H correlation spectroscopy (COSY; Aue et al., 1976), or total correlation spectroscopy (TOCSY; Braunschweiler and Ernst, 1983; Bax and Davis, 1985), and ^{1}H-^{1}H 2D NOESY spectra. For larger molecules, the overlap between NMR signals complicates the assignment. Such systems require the addition of a third dimension such that the resonances are also spread along the new dimension, reducing the overlap. Typically, this is combined with the replacement of the naturally abundant isotope ^{14}N with ^{15}N, which is more suitable for NMR studies, and the replacement of the NMR inactive ^{12}C nuclei with its NMR active isotope ^{13}C. This allows to further decrease overlap and to extract additional chemical shift information from these nuclei. In these cases, assignment is based on triple resonance experiments, which allow sequential assignment of the backbone nuclei, combined with experiments that use a TOCSY transfer to assign the side chain atoms (Cavanagh et al., 1996).

In case of extreme overlap, standard strategies might not suffice to obtain complete assignments. For instance, in my study of the complex between the extended Mad1-SID and the PAH2 domain, the assignment of the SID was heavily complicated by spectral overlap. The conventional assignment strategy in the study of protein complexes is to use 2D homonuclear proton spectra for the smaller, unlabelled domain, in this case the SID. To overcome the overlap in the 2D spectra, I developed a 3D homonuclear TOCSY-TOCSY experiment (discussed in Chapter 3).

After completion of the assignment, structural restraints can be derived. The main sources of structural restraints in protein structure determination are the interproton distances derived from NOESY experiments, orientational restraints derived from RDC measurements and dihedral restraints derived from chemical shifts or *J*-couplings. These experimental restraints, have to be combined with *a priori* knowledge about the covalent structure and the stereochemistry of the amino acids to determine the position of all atoms in the molecule. In short, special search algorithms are used that can simultaneously optimize the agreement of random starting structures with all experimental restraints and the *a priori* knowledge (reviewed in Clore and Gronenborn, 1998). The resulting set of structures can be further refined by specialized protocols (Spronk et al., 2002; Linge et al., 2003). Finally, the quality of the resulting structures needs to be validated (reviewed in Spronk et al., 2004).

Protein dynamics by NMR

Global and local molecular motions causes random, time dependent fluctuations of the Hamiltonian that can be effective as a relaxation mechanism for the nuclear spin. Relaxation parameters thus reflect the overall and local dynamics of the molecule. However, the number of relaxation parameters that can be measured is not sufficient to allow a complete reconstruction of the underlying motions (Lipari and Szabo, 1982a). Furthermore, analysis of these parameters can be complicated as the different spins in a molecule can form a dense, coupled network of spins. For this reason, study of protein dynamics generally focuses on the ^{15}N backbone nucleus, which is mainly relaxed by its chemical shift interaction and the dipolar interaction with its directly attached proton. This allows a relatively simple interpretation of its relaxation parameters. Typically, three relaxation parameters are measured: the longitudinal relaxation rate R_1, responsible for reestablishing thermal equilibrium; the transverse relaxation rate R_2, responsible for the loss of phase coherence of the NMR signal; and the steady state nuclear Overhauser effect NOE between the ^{15}N nucleus and its attached proton. The ratio R_2/R_1 reflects mostly the global rotational diffusion and can be used to estimate the rotational correlation time or to estimate the overall shape of the molecule. The NOE reflects mostly local internal motions of the N-H bond on a picosecond to nanosecond time scale. A more detailed and quantitative description of the mobility of the N-H vector can be obtained from the combined analysis of the ^{15}N R_1, R_2 and the {^{1}H}-^{15}N NOE using spectral density mapping (Farrow et al., 1995; Lefèvre et al., 1995; Ishima and Nagayama, 1996) or the model-free approach (Lipari and Szabo, 1982ab; Clore et al., 1990). A more extensive characterization of protein dynamics is possible by measuring relaxation parameters of different nuclei either in the backbone or the side chain.

NMR is not only sensitive to fast molecular motions occurring on pico- to nanosecond time scale, but also slower motions can be studied (see also Figure 1.4). Notably, motions on a microsecond to millisecond time scale are thought to be important for biological processes such as ligand binding, enzyme catalysis, allosterics and protein folding and have received considerable attention during the past few years (Mulder et al., 2001; Volkman et al., 2001; Eissenmesser et al., 2002; Korzhnev et al., 2004; Dyson & Wright, 2005). Such motions will change the chemical environment of the nuclear spin and thus cause a fluctuation of the chemical shift on a µs-ms time scale. These fluctuations cannot induce transitions between nuclear spin states but result in an accelerated loss of phase coherence and thereby increase the effective transverse relaxation rate R_2. In general, any process whereby a spin experiences different chemical environments on a µs-ms time scale, will result in an increased R_2 relaxation rate. When this process is caused by internal motions in a molecule, it is called conformational exchange, whereas the term chemical exchange refers to a spin exchanged between different molecules by a chemical process. These exchange processes can occur between any number of sites. In the simplest case, exchange can be described as a two-site process between a state *A* and state *B*:

$$A \underset{k_{BA}}{\overset{k_{AB}}{\rightleftarrows}} B \qquad [1.3]$$

where k_{AB} is the forward rate constant and k_{BA} the backward rate constant (in s^{-1}). Exchange between the two states is assumed to occur instantaneously compared to the average life times of state *A*, $\tau_A = 1/k_{AB}$, and state *B*, $\tau_B = 1/k_{BA}$. The equilibrium constant *K* defines the relative populations of state *A* (p_A) and state *B* (p_B):

$$K = \frac{p_B}{p_A} = \frac{k_{BA}}{k_{AB}} \qquad [1.4]$$

The time evolution of a transverse spin coherence *V*, assuming that it is not coupled to any other coherence, is given by the modified Bloch equations (McConnell, 1958):

$$\frac{d}{dt}\begin{bmatrix} V_A(t) \\ V_B(t) \end{bmatrix} = \begin{bmatrix} -i\omega_A - k_{AB} - R_A & +k_{BA} \\ +k_{AB} & -i\omega_B - k_{BA} - R_B \end{bmatrix} \begin{bmatrix} V_A(t) \\ V_B(t) \end{bmatrix} \qquad [1.5]$$

where V_i, ω_i and R_i are the amount of coherence *V*, its resonance frequency (in rad·s^{-1}) and its relaxation rate in state *i* (in s^{-1}). The solution of this coupled differential equation shows that the NMR signals corresponding to V_A and V_B will have both an exchange induced additional line broadening (R_{ex}) and an exchange induced shift of the observed resonance frequency (ω_{ex}) (Woessner, 1961). The amount of line broadening and frequency shift is very sensitive to the *exchange regime*, defined by the ratio of the exchange rate constant k_{ex} (equal to $k_{AB} + k_{BA}$) and the frequency separation between the two states $|\Delta\omega|$ (defined by $|\omega_A - \omega_B|$) (see Figure 1.5).

Many biological processes involve asymmetric exchange, where the relative populations of the two states are skewed and one state is dominant (e.g. $p_A > p_B$). In this case, the signal of the minor state can be unobservable even in slow exchange conditions, because of the greater R_{ex} contribution to its line width (see also Figure 1.5). A quantitative description of the underlying exchange process can still be obtained using so-called relaxation dispersion experiments, where the transverse relaxation rate of the major state coherence V_A is measured as a function of interpulse delay between the refocusing 180° pulses in Carr-Purcell-Meiboom-Gill (CPMG) experiments (reviewed in Palmer et al., 2001). Decreasing the interpulse delay reduces the amount of dephasing due to exchange, resulting in lower relaxation rates. The dependence of the relaxation rate on the pulsing rate is approximated for all exchange regimes by the Carver-Richards-Jen equation (Carver and Richards, 1972; Jen, 1978). In favorable cases, fitting of the experimental rates to this equation yields kinetic and thermodynamics information in form of the exchange rate constant and the populations of the two states.

In addition, analysis of relaxation dispersion experiments yields the magnitude of the frequency separation between the two states. This information is very useful to character-

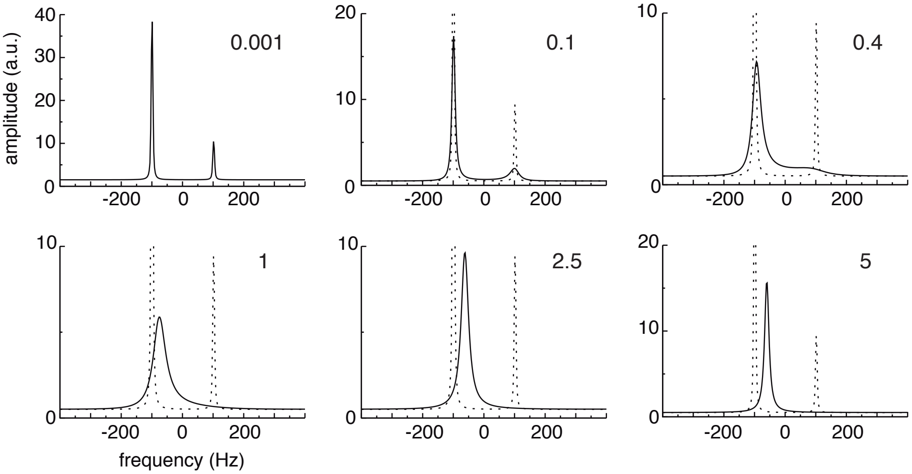


Figure 1.5. Simulated line shapes for asymmetric exchange (p_A = 0.8, p_B = 0.2; $\Delta\omega$ = 2π·200 rad·s^{-1}) for different exchange regimes as indicated by the ratio $k_{ex}/\Delta\omega$, shown in the right upper corner of each plot. The first plot in the top row shows the spectrum in ultimately slow exchange ($k_{ex}/\Delta\omega$ = 0.001), in which the expected relative intensities for the two states are observed. An increased exchange rate ($k_{ex}/\Delta\omega$ = 0.1) results in more severe line broadening of the minor state. The plots at $k_{ex}/\Delta\omega$ = 0.4 and 1 are in the intermediate exchange regime where line broadening is maximal and the line shape appears distorted (note the differences in amplitude). After coalescence (at $k_{ex}/\Delta\omega = 1/\sqrt{2}$), the minor resonance cannot be distinguished anymore, while the exchange induced shift of the major resonance is easily observed. In fast exchange ($k_{ex}/\Delta\omega$ >> 1) a single peak is observed at the population weighted average position, of which the line width is reduced by motional narrowing. The spectrum of ultimately slow exchange is shown with dashed lines in each plot for comparison.

ize the structure of the unobservable minor state, provided the sign of the frequency separation is known. Recently, methods have been proposed to measure this sign for single quantum ^{15}N coherence (Skrynnikov et. al., 2002; Korzhnev et al., 2005). In Chapter 6, I present a new method to measure the sign of the frequency difference between the major and minor state that, in contrast to previously proposed experiments, is not limited to ^{15}N but can be applied to any nuclear spin. Application of this method to the unbound PAH2 domain led to the structural characterization of a minor conformation with a relative population of ca. 1%.

1.4 Outline thesis

In this chapter, I have presented a general overview of the biological role of the Sin3 protein and the role of its PAH domains. I have discussed the role of NMR in the study of the relation between protein structure, dynamics and function and I have introduced some basic NMR theory that I used to develop new NMR experiments. Chapter 2 describes some unexpected artifacts caused by decoupling and presents a simple solution to prevent this artifact. Chapter 3 introduces a new NMR experiment for the assignment of complex, unlabeled molecules bound to a labeled molecule. The 3D X-filtered TOCSY-TOCSY experiment reduces spectral overlap and allows the assignment of the extended SID of Mad1 complexed to the PAH2 domain of Sin3. In Chapter 4, the structure of this complex is presented and its implications for the interaction between Mad1 and Sin3 is discussed. Chapter 5 is focused on the unbound state of the PAH2 domain and discusses the role of structural and dynamical plasticity in the interaction with Sin3 interacting domains from repressor proteins.

In Chapter 6, a new method to characterize the structure of unobservable protein conformations is proposed. Finally, Chapter 7 presents a general discussion of the results.

Chapter 2.

Hugo van Ingen
Geerten W. Vuister
and Marco Tessari

Journal of
Magnetic Resonance
(2002), **156**, p. 258-261

A two-dimensional artifact from asynchronous decoupling

Acknowledgements. We would like to thank Prof. dr. C.W. Hilbers for his stimulating interest and Tine Walma for her constructive comments on the manuscript. This research is financially supported by the Netherlands Organization for Scientific Research, NWO.

Many heteronuclear NMR experiments employ decoupling to collapse the heteronuclear multiplet, using decoupling schemes with a periodic phase modulation like WALTZ, MLEV or GARP. Because of the periodic nature of these schemes, cycling sidebands are generated, whose intensity can be strongly reduced by decoupling asynchronously. We show that the most common implementation of asynchronous decoupling on modern spectrometers is such that the cycling sidebands are subjected to a periodic modulation. For multidimensional experiments, this results in ridges that can seriously compromise the quality of the spectrum. Based on our model, the artifact in a 2D $\{^1H\}$-^{15}N NOE equilibrium experiment are simulated and it is shown that the artifact can be prevented by using synchronous decoupling.

Introduction

Heteronuclear NMR experiments on biomolecular systems often require decoupling during acquisition to collapse the multiplet resulting from scalar coupled spins. Heteronuclear decoupling removes the effects of the scalar coupling and typically yields both a higher resolution and sensitivity. Phase modulated schemes like WALTZ, MLEV and GARP (Levitt et al., 1982; Shaka et al., 1983a; Shaka et al., 1985) have greatly improved the quality of decoupling, allowing for efficient removal of the scalar coupling at relatively low power, while maintaining a large decoupling bandwidth. The drawback of these sequences is that they generate so-called *cycling sidebands* because data acquisition occurs at intervals other than the refocusing of the *J*-coupling (Freeman et al., 1982; Shaka et al., 1983b; Shaka et al., 1986). Cycling sidebands are undesirable because they can overlap with signals of interest and reduce the intensity of the parent signal (Shaka et al., 1983b; Kupče et al., 1996). The intensity of these sidebands can be strongly attenuated by decoupling asynchronously in combination with time-averaging (Freeman et al., 1982; Shaka et al., 1983b; Shaka et al., 1986).

In the asynchronous mode, decoupling starts from a different point in the phase modulation pattern for each acquisition, resulting in a (partial) randomization of the phase of the sidebands (Shaka et al., 1983b; Shaka et al., 1986). In contrast, in the synchronous mode, decoupling starts each time from the beginning of the modulation sequence, resulting in identical sidebands. For example, asynchronization can be achieved by changing the decoupler pulse length (Dijkstra, 1989) or by cyclically permuting an element of the phase modulation sequence (Shaka et al., 1986). The most commonly used method on modern spectrometers is depicted in Figure 2.1, which we refer to as the "jump-resume" implementation.[†] In this implementation decoupling is stopped at the end of acquisition, potentially in the middle of an element of the phase modulation scheme. In the successive scan, at the start of acquisition, decoupling is resumed from the following element in the decoupling sequence.

Here we show that, for a multidimensional experiment, this implementation of asynchronous decoupling introduces a periodic modulation of the cycling sidebands, resulting in a multidimensional artifact with defined frequencies in the indirect dimension. As an example, Figure 2a shows the spectrum of a 2D heteronuclear NOE experiment employing asynchronous WALTZ-16 decoupling. The quality of the spectrum is seriously compromised by several horizontal ridges, making it impossible to determine the peak intensity for a number of residues. We show that these ridges, while they appear to be modulated in the indirect dimension, are actually generated by the cycling sidebands in the direct dimension. Hence, we call this a two-dimensional artifact.

Theory

Ideally, asynchronous decoupling would involve complete randomization of the decoupling starting point to achieve optimal suppression of sidebands, since their phase and intensity depend on the point in the modulation pattern at which decoupling is started (Shaka et al., 1986). However, using a "jump-resume" implementation of asynchronous decoupling, the starting point is not randomized. Although from scan to scan decoupling starts from different elements in the sequence, a periodicity is generated equal at most to the number of elements in the phase modulation sequence. The number of elements is different for different decoupling schemes and depends on their specific implementation.

Since our experiments employed WALTZ decoupling, we shall now extend our arguments for this case, but similar reasoning holds for other decoupling schemes as well. Figure 2.1 schematically depicts the implementation of WALTZ-16 decoupling (Shaka et al, 1983a). Although its bandwidth is relatively limited, WALTZ-16 generates few sidebands, making it an appropriate choice for ^{15}N-decoupling. The implementation of

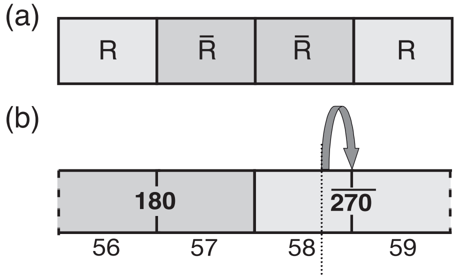


Figure 2.1. (a) Schematic representation of the WALTZ-16 supercycle (Shaka et al., 1983a), where R is given by 270-$\overline{360}$-180-$\overline{270}$-90-$\overline{180}$-360-$\overline{180}$-270, the numbers refer to the pulse rotation angles and the overbars indicate phase-inverted pulses. The total length of the supercycle is 96 times the 90°-decoupler pulse length. (b) Implementation of "jump-resume"-asynchronous decoupling. The boxes represent the 90°-elements of the modulation pattern, the numbers in the boxes refer to the elements of R, the numbers below the boxes indicating the index of that 90°-element. When acquisition is stopped somewhere in a 90°-element (indicated by the dotted line), decoupling resumes from the next 90°-element at the following acquisition, this "jump" is indicated by the arrow. This figure depicts the situation when decoupling started at element 1 in the previous scan, using Equation [2] with an acquisition time equal to 0.0768 sec and a 90° decoupling pulse of 222 μs.

[†] We have observed this artifact on different types of spectrometers of different manufacturers. Personal communication with the technical departments of both Bruker A.G. and Varian Inc. confirmed the "jump-resume" implementation of asynchronous decoupling.

WALTZ-16 decomposes the modulation sequence, or supercycle, in 96 elements, each of the duration of a 90°-pulse at decoupling power and each with a certain phase (cf. Figure 2.1b). Therefore, decoupling starts from the same element again after at most 96 scans, generating identical sidebands in scan 1 and 97, 2 and 98 and so on. The exact period can be calculated, considering that the starting element is shifted regularly from scan to scan by a fixed number of 90°-elements, solely depending on the durations of the acquisition time and the WALTZ-16 supercycle. This shift can be expressed as a fraction of a supercycle, as indicated by the following expressions:

$$shift = \frac{n}{96} \qquad [2.1]$$

where

$$n = \mathbf{int}\left(96 \times \mathbf{fract}\left(\frac{at}{W16}\right)\right) + 1 \qquad [2.2]$$

and *at*, *W16* denote acquisition time and duration of the WALTZ-16 supercycle respectively; **int**(x) returns the integer part of x, **fract**(x) returns the decimal part of x, and n is the number of 90°-elements by which the starting element is shifted each scan (cf. Figure 2.1).

From Equation [2.1] it can be deduced that, independent of *at* and *W16*, the shift is equal to an integer multiple of supercycles after at most 96 scans, but smaller periods can also be expected for some values of n. As a result, the sidebands are modulated periodically and consequently cancellation artifacts, resulting from pairwise subtraction of scans, will be modulated periodically too. According to the Fourier theorem, the cancellation artifact can then be described as a superposition of sinusoidal components oscillating at the fundamental frequency and its harmonics.

For a multidimensional experiment, the cycling sidebands and the cancellation artifact accumulated in every FID are modulated with a certain period of FIDs, depending on the number of scans per FID. The periodicity of this modulation results, after Fourier transformation, in horizontal ridges through the spectrum at a specific positions in the indirect dimension given by the fundamental frequency and its harmonics.

Results and discussion

The heteronuclear {^{1}H}-^{15}N NOE experiment (cf. Figure 2.3a) is an example of a multidimensional experiment which, due to its low sensitivity, may suffer from the artifact described above. Figure 2.2a shows the equilibrium spectrum of the {^{1}H}-^{15}N NOE experiment performed on a uniformly ^{13}C/^{15}N-labeled 12.5 kDa protein studied in our laboratory. The artifact shown is common for any NOE equilibrium experiment run on a modern spectrometer using asynchronous decoupling and a pulse sequence relying on phase cycling to remove the equilibrium proton magnetization. Since the latter magnetization is 10 times larger than the signal originating from the nitrogen spins, the ridges generated by the cycling sidebands of the equilibrium proton signals are relatively large compared to the selected signals.

The position of the ridges can be rationalized using the following equation:

$$ns_{cp} \times shift = shift_{cp} \qquad [2.3]$$

where ns_{cp} is the number of scans per complex point, *shift* is given by Equation [2.1], and $shift_{cp}$ is the shift of the starting element per complex point.

Substitution of the experimental settings for *at* and *W16* (76.8 ms and 96*222 μs respectively, cf. Figure 2.2) in Equation [2.2] shows that the starting point of decoupling is shifted by 58 90°-elements for each transient (n), resulting, after substitution in Equation

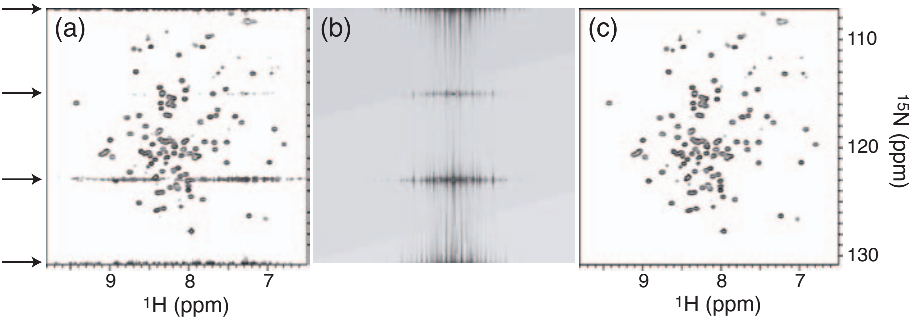


Figure 2.2. (a) The {^{1}H}-^{15}N NOE equilibrium spectrum of the PAH2:MAD-SID24 complex using asynchronous WALTZ-16 decoupling with 1.12 kHz RF-field strength (90°-pulse of 222 μs). The arrows indicate the ridges. The sample was dissolved in an 95% H_2O/5% D_2O mixture, the pH was set to 6.3 using a phosphate buffer and 50 mM of KCl as well as trace amounts of NaN_3 and protease inhibitor were added. Experimental parameters: acquisition time was set to 0.0768 sec, the spectral width was 10000 Hz (t_2) x 1200 Hz (t_1), 32 scans were averaged per FID and 120 complex points were recorded. (b): Simulated spectrum of an AX spin-system with asynchronous WALTZ-16 decoupling, using the experimental parameters of (a). (c) Same as (a), but now with synchronous decoupling. All spectra were acquired on a Varian Unity Inova 500 MHz spectrometer. Spectra were processed using NMRPipe (Delaglio et al., 1995).

[2.1], in a *shift* per scan equal to 58/96 supercycle. Because 32 transients were averaged per FID and States-TPPI quadrature detection (Marion et al., 1989) was used, ns_{cp} is equal to 64, resulting, after substitution in Equation [2.3], in a shift per complex point equal to 38. As a consequence, after three complex points, the shift is equal to an integer number of supercycles. Thus, the cancellation artifact is modulated with a period of three complex points, resulting in two ridges at a frequency of +/- one-third of the spectral width in the indirect dimension.

Figure 2.2a shows two ridges along the edges of the spectrum and two ridges dividing the total spectrum into three equal parts, as indicated by the arrows. Since States-TPPI was used to shift any signal not modulating in t_1 by half the sweep width, the zero frequency reference for the artifact is at the border of the spectrum. In perfect agreement with the theoretical period of three complex points, two ridges appear at a frequency of +/- one-third of the sweep width of the indirect dimension (i.e. +/- one-sixth of the sweep width relative to the center of the spectrum). The ridges at the border stem from the zero-frequency component of the artifact and/or aliased harmonics.

To simulate this two-dimensional artifact, a program was written in OCTAVE (numerical computation software freely available from http://www.octave.org.), which simulates the effect for an *AX* two-spin system. After generating $2A_yX_z$ antiphase coherence, spin *X* is subjected to asynchronous decoupling using a WALTZ-16 modulation scheme (cf. Figure 2.3b). The evolution of the density matrix of this system is calculated numerically, taking Larmor precession, *J*-coupling, relaxation and decoupling into account. The receiver phase is inverted every second scan, thus allowing complete focus on the cancellation artifact generated by the cycling sidebands of the main signal. This basic 1D-experiment was expanded into a 2D-experiment by repetitive evaluation using the experimental parameters of the actual {^{1}H}-^{15}N NOE experiment. The final calculated two-dimensional data matrix was then processed in an identical fashion as the experimental data.

The resulting spectrum displays ridges at positions identical to those observed in the experimental spectrum (cf. Figure 2.2b). In

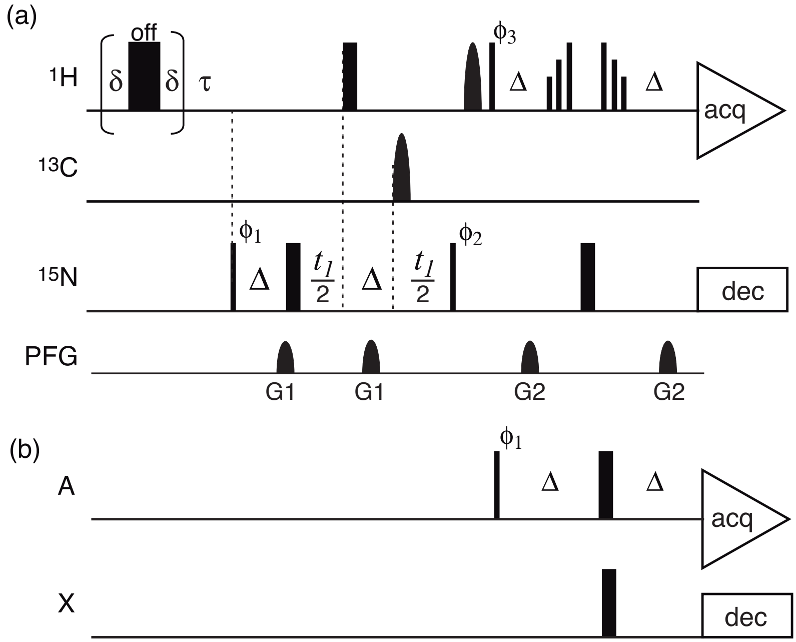


Figure 2.3. (a) Pulse sequence of the {^{1}H}-^{15}N NOE experiment. The saturation pulse train composed of 135°-pulses is given off-resonance to record the equilibrium spectrum. The delay Δ is set to ¼ $J_{\mathrm{N-H}}$. The following phase cycle is used to cancel the equilibrium proton magnetization and to select the magnetization originating from the nitrogen spins: $\phi_1 = x,-x$; $\phi_2 = y,y,-y,-y$; $\phi_3 = -x$; $\phi_{\mathrm{rec}} = x,-x,-x,x$. States-TPPI was implemented on ϕ_1 (Marion et al., 1989). The water was put along the z-axis using a half-gauss shaped water flip-back pulse. The residual water in the x,y-plane was suppressed by 3-9-19 WATERGATE pulses (Sklenar et al., 1993) in the reverse INEPT-block. Gradient strengths were 24 and 15 Gauss/cm for G1 and G2 respectively. (b) Pulse sequence used in the simulation program to calculate Figure 2b. Spin X is decoupled asynchronously using WALTZ-16 modulation with a "jump-resume" implementation. The phase cycle $\phi_1 = x,x$; $\phi_{\mathrm{rec}} = x,-x$ cancels the equilibrium magnetization of spin A. The delay Δ is set to ¼ $J_{\mathrm{A-X}}$.

addition, similar relative intensities are reproduced. Figure 2.2b also clearly shows the two-dimensionality of the artifact. A single signal generates many cycling sidebands appearing at the decoupler cycling rate and its harmonics (Freeman et al., 1982), and all are subjected to identical modulations, thus creating ridges parallel to the horizontal axis. While other studies describing artifacts from cycling sidebands in two dimensions (Zhang and Gorenstein, 2000; Kupče and Freeman, 2001) have used decoupling in the indirect dimension, we show that the artifact introduced by asynchronous decoupling during t_2 is inherently two-dimensional.

Since the artifact is caused by the cycling sidebands of the large unwanted equilibrium proton signal, it can be prevented by decoupling synchronously. Each scan generates identical sidebands ensuring perfect subtraction of the unwanted signal. This is shown in Figure 2.2c, which displays the {^{1}H}-^{15}N NOE experiment recorded with synchronous decoupling, but using otherwise identical experimental parameters as for Figure 2.2a. As is immediately evident, the artifact has completely disappeared.

Alternatively, the artifact can be prevented by using a different implementation of asynchronous decoupling, ensuring a truly random starting element of the decoupling sequence. This would lead to the loss of the modulation of the cycling sidebands, resulting in a slight increase of t_1-noise, as confirmed by our simulations (data not shown).

Conclusion

We have shown that the “jump-resume” implementation of asynchronous decoupling results in a periodic modulation of the cycling sidebands. For a multidimensional experiment this results in ridges through the spectrum, as shown for the heteronuclear NOE equilibrium experiment. This artifact is multidimensional in nature, since the apparent frequencies along the indirect dimension actually stem from modulations in the direct dimension. Other phase modulated decoupling schemes like GARP and MLEV have been shown to give rise to similar artifacts (data not shown).

The artifact presented here, is a common problem for any insensitive experiment that relies on phase cycling to cancel a relatively large unwanted signal and uses asynchronous decoupling. Synchronous decoupling and a different implementation of asynchronous decoupling are possible solutions to this problem.

Chapter 3.

Hugo van Ingen,
Marco Tessari
and Geerten W. Vuister

*Journal of
Biomolecular NMR*
(2002), **24**, p. 155-160

A 3D doubly sensitivity enhanced X-filtered TOCSY-TOCSY experiment

Acknowledgements. We would like to thank Prof. Dr. C.W. Hilbers for his stimulating interest and Tine Walma for her constructive comments on the manuscript. This research is financially supported by the Netherlands Organization for Scientific Research, NWO.

We present a 3D double sensitivity enhanced X-filtered homonuclear TOCSY-TOCSY experiment for the assignment of unlabeled molecules complexed to labeled protein- or nucleic acid-domains. The resulting spectrum is clean, can be measured in a reasonable amount of time and allows for increased resolution of overlapping resonances when compared to 2D methods. The 3D X-filtered TOCSY-TOCSY allows for assignment in cases where the size or the composition of the unlabeled molecule results in a high degree of overlap.

Introduction

Understanding of biochemical processes requires knowledge of interactions between biomolecules, whether they are proteins, nucleic acids or drug molecules. Structural studies of complexes between interacting biomolecules provide valuable information about the structural details of molecular recognition, binding mechanisms, binding specificity and can assist the design of small molecule ligands for proteins or nucleic acids (Zhang and Yuan, 1998; Lo Conte et. al., 1999; Nadassy et. al., 1999; Dongen et. al., 2002). NMR has been established as a technique that can provide such a detailed analysis of these interactions at atomic level (Ikura et. al., 1992; Theriault et. al., 1993; Omichinski et. al., 1993; Billeter et. al., 1993; Chuprina et. al., 1994; Qin et. al., 2001; Pochapsky and Pochapsky, 2001; Zuiderweg, 2002). Typically in these studies, one of the molecules is enriched (labeled) with ^{13}C and/or ^{15}N isotopes, while the other molecule is not. Assignment of the unlabeled molecule is usually performed using the conventional sequential-assignment approach (Wüthrich, 1986), relying on homonuclear 2D X-filtered NOESY and TOCSY spectra (Otting et. al., 1986; Ikura and Bax, 1992).

There are cases, however, in which the classical 2D approach falls short, because the spectra show too many overlapping resonances. This problem can be overcome by swapping the isotope labeling of the molecules in the complex. Often, however, this procedure might be too costly or simply not practical, particularly when the ligand is obtainable only by direct isolation from a natural source or cannot be produced effectively by biochemical methods. Here, we report an efficient alternative, the use of a homonuclear 3D double sensitivity enhanced X-filtered TOCSY-TOCSY. We have used this experiment to obtain the complete ^{1}H assignment of an extended fragment of the Sin Interacting Domain (SID) of the protein Mad1 complexed to the PAH2-domain of the protein Sin3B. In contrast to a previous study by Spronk et. al. (2000) where a minimal SID-fragment of 13 residues was used, the conventional 2D methods failed on the extended fragment. This extended SID consists of 24 residues: a three residue extension to the N-terminus with respect to the minimal SID and eight additional residues at the more flexible C-terminus, of which 6 residues are either arginines or glutamic acids. This redundancy in amino acid composition resulted in extensive overlap in the 2D X-filtered TOCSY spectrum. However, combining information from the 3D X-filtered TOCSY-TOCSY spectrum and a 2D X-filtered NOESY spectrum resulted in full assignment of the extended SID fragment with exception of the N-terminal residue.

Pulse sequence design

Previously, 3D homonuclear TOCSY-TOCSY experiments have been developed for the assignment of unlabeled proteins or oligosaccharides (Cieslar et. al., 1990; Rutherford and Homans, 1995). These experiments, however, have not gained much popularity since the relatively low transfer efficiency of a TOCSY experiment results in a low sensitivity. Furthermore, sensitivity is decreased in a regular multi-dimensional TOCSY experiment because the signal resulting from one of the two orthogonal operators I_x and I_y is usually lost by the use of a non-isotropic mixing sequence and/or the use of trim pulses or z-filtration in order to remove spectral artifacts. In addition, the number of peaks in a homonuclear experiment increases largely in going from 2D to 3D as a result of the presence of an additional mixing period. (Vuister et. al., 1990). In the present study, we used short TOCSY mixing times for both transfer periods, limiting the number of cross peaks. In addition, we included a ^{13}C/^{15}N filter to remove signals from the labeled molecule in the complex and we maximized the sensitivity by incorporating sensitivity enhancement schemes for both indirect dimensions.

Both equivalent orthogonal pathways can be retained in a TOCSY-experiment using an isotropic mixing sequence and sensitivity enhancement (SE) (Cavanagh and Rance, 1990; Krishnamurthy, 1995; Sattler et. al., 1995; Wijmenga et. al., 1996; Köver et. al., 1998). In the present implementation, we incorporated a double sensitivity enhancement scheme and thus the 3D experiment retains full sensitivity and is up to twice as sensitive as a regular 3D (Krishnamurthy, 1995; Sattler et. al., 1995). Importantly, the insertion of SE in this homonuclear X-filtered experiment did not require additional pulses or delays.

The pulse field gradient coherence selection method (Kay et. al., 1992) was used to implement the double sensitivity enhancement in a similar way as previously outlined by Wijmenga et. al. (1996) (cf. Figure 3.1). The magnetization is dephased in both indirect dimensions, t_1 and t_2, by self-compensating gradients G1 and G2, respectively. Since the gradients are inserted in the X-filter delays, there are no additional relaxation losses due to the inclusion of the SE. Gradient G3 is used to refocus the observable magnetization by the echo-antiecho approach and is combined with the WATERGATE gradient G0 (Piotto et. al., 1992) and to refocus the observable magnetization, following the echo-antiecho approach. The phases ϕ_2 and ϕ_4 are cycled such to generate PP-, PN-, NP-, or NN-type FIDs. Adding and subtracting of the appropriate FIDs results in a final hypercomplex data set that can be Fourier trans-

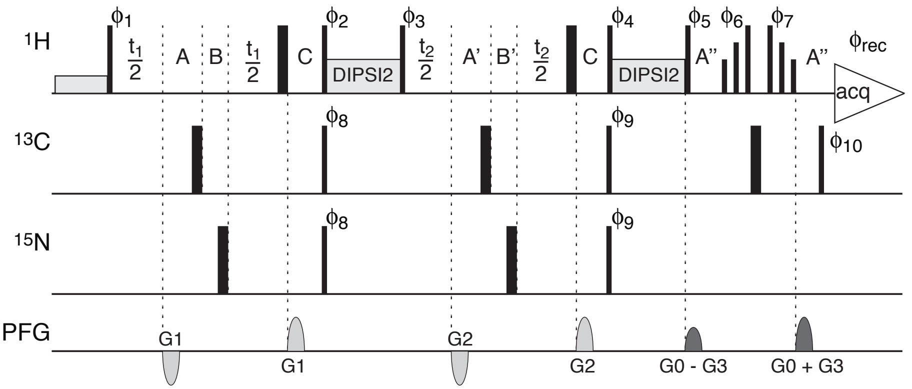


Figure 3.1. Pulse sequence of the 3D double SE-PFG coherence selection $^{13}C/^{15}N$-filtered TOCSY-TOCSY. Thin and thick bars represent high-power 90°- and 180°-pulses for the proton, carbon and nitrogen-channel at 33.4, 17.1 and 6.6 kHz RF-field strength, respectively. For optimal water suppression we used a 3-9-19 WATERGATE at constant power (Piotto et. al., 1992) and low power (38 Hz RF-field strength) presaturation. The filter delays A, A' and A'' were set to match $1/(4J_{CH})$ for the methyl, the aliphatic and the aromatic proton-carbon couplings (2, 1.79 and 1.5 ms, respectively). The ^{13}C carrier is placed at 40 ppm for filter delays A and A' and is shifted to 130 ppm prior to delay A''. Delay C is set to $1/(4J_{NH})$ (2.75 ms) and delay B and B' were set to 0.75 ms and 0.96 ms in order to match the sum of A/A', B/B' and C to $1/(2J_{NH})$. The DIPSI-2 sequence was used for the isotropic TOCSY transfer (Shaka et. al., 1988). The DIPSI mixing lock axis was set orthogonal to ϕ_2 and ϕ_4 and the mixing time was set to 21 ms for both mixing periods with an RF-field strength of 9.7 kHz. All gradients $Gn(x, y, z)$ were square-shaped, given along three axis, and were set to the following strengths in Gauss/cm: G1(-7.5, 0, 15); G2(0, -9.3, 18.6); G0(1.5, 0, 0). Gradient duration was set to 0.5 ms for all gradients. G0 is a WATERGATE gradient and is combined with the refocusing gradient G3. The strength of gradient G3 is controlled by the following equation: $G3 = s_1 \cdot G1 + s_2 \cdot G2$, in which s_1 and s_2 are either +1 or -1. PP, NP, PN, NN-type FIDs are obtained by a 180° phase shift on ϕ_2 and inversion of s_1 for every second point in t_1 and a 180° phase shift on ϕ_2 and ϕ_4 and inversion of s_2 for every second point in t_2 (cf. Table 1). All pulses are given along the x-axis unless specified otherwise: $\phi_1 = 2(y, -y, x, -x)$; ϕ_2, ϕ_3, $\phi_5 = 2(x, x, y, y)$; $\phi_4 = 2(-x, -x, -y, -y)$; $\phi_6 = 4(x), 4(y)$; $\phi_7 = 4(-x), 4(-y)$; $\phi_8 = 4(x), (-x)$; $\phi_9 = 4(x), 4(-x)$; $\phi_{10} = 4(x, -x)$; $\phi_{rec} = x, -x, -y, y, -x, x, y, -y$. In the acquisition dimension, the spectral width was set to 9000 Hz and 768 complex points were acquired. In both indirect dimensions, 122 echo-antiecho pairs were acquired sampling a 6000 Hz spectral width.

formed to a 3D spectrum with in-phase line shapes and improved sensitivity.

TOCSY experiments are not usually run using sensitivity enhancement, since this reinforces phase-twisted line shapes. This effect originates from the transfer of in-phase to anti-phase magnetization, which is inherent for an isotropic mixing sequence. For larger molecules, however, this is less of a problem, because the larger linewidths result in self-cancellation of the anti-phase components. In the present study, the unlabeled molecule of interest is a protein fragment tightly bound ($K_D = 0.2$ μM) to a protein domain in a 1:1 complex of 12.5 kDa. Consequently, the linewidths of the peptide signals are as large as those of the protein, resulting in a spectrum virtually free from interfering artifacts. Furthermore, excellent ^{13}C- and ^{15}N-filtering is obtained via three independent ^{13}C-filter delays and two ^{15}N-filter delays. The three ^{13}C-filter delays were set to match the methyl, aliphatic and aromatic $^1J_{CH}$ coupling constant.

Results

Data were acquired at 20.0°C on a Varian 600 MHz Unity Inova spectrometer equipped with a triple-resonance triple-axis gradient probe. The sample contained 1.3 mM 1:1 complex of unlabeled hMad1-SID24 bound to uniformly $^{13}C/^{15}N$ labeled mSin3B-PAH2 at pH 6.3 and was prepared as described before (Spronk et. al., 2001). Using 8 scans per FID, 122x122 echo-antiecho pairs were recorded (~5 days of measuring time) to sample the complete proton spectrum in both indirect dimensions. Reducing the spectral width in these dimensions can decrease the total measurement time by allowing for folding of the spectrum. The four successive PP, PN, NP and NN-type FIDs were recorded as shown in Table 3.1 and were combined as outlined by Wijmenga et. al. (1996). The data were processed in straightforward way using NMRPipe (Delaglio et. al., 1995).

Table 3.1
Phase cycling scheme and gradient signs used to generate echo-antiecho pairs.[a]

FID	type (t_1,t_2)	ϕ_2	ϕ_4	s_1	s_2
1	PP	$+x$	$-x$	+1	+1
2	NP	$-x$	$-x$	−1	+1
3	PN	$-x$	$+x$	+1	−1
4	NN	$+x$	$+x$	−1	−1

[a] The phases ϕ_2 and ϕ_4 refer to the 90°-pulses prior to the two TOCSY mixing periods. The coefficients s_1 and s_2 are used to control the strength of refocusing gradient G3, G3 = $s_1 \cdot G1 + s_2 \cdot G2$. See also the caption to Figure 3.1

The increased resolution of the 3D spectrum made it possible to obtain a complete and specific assignment of the 24-residue peptide (N-VRMNIQMLLEAADYLERREREAEH-C). Figure 3.2 illustrates this. The outer left panel (a) plots the region from the spectrum of the conventional 2D X-filtered TOCSY experiment where contacts are observed between the H_α (F1-axis) and the H_β and H_γ (F2-axis) of four overlapping glutamic acids, E20, E23, E25 and E27. Three of the four glutamic acids can be identified in the 2D spectrum. The H_α frequencies of E23 and E25 can be separated in the 2D, but E20 and E27 have nearly identical chemical shifts for their H_α. The 3D spectrum of the doubly sensitivity enhanced X-filtered TOCSY-TOCSY resolves these two spin systems and also resolves the $H_{\gamma 3}$ of E25 and $H_{\gamma 2}$ of E27. This is shown in the next three panels showing corresponding sections of the (F2, F3)-planes of the 3D

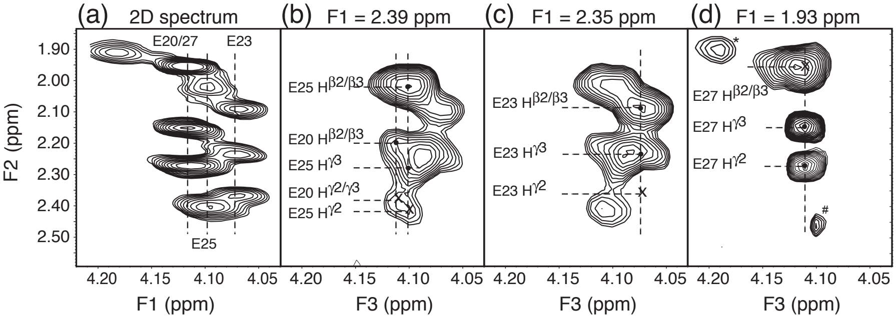


Figure 3.2. (a) section from the spectrum of a 2D $^{13}C/^{15}N$-filtered sensitivity enhanced TOCSY experiment showing the contacts between the H_α and H_β and H_γ for the four overlapping glutamic acids E20, E23, E25 and E27 of the SID24-peptide. The three dashed vertical lines indicate the H_α frequencies of these residues; the labels refer to their residue numbers. (b), (c) and (d) are sections of (F2, F3)-planes of the 3D $^{13}C/^{15}N$-filtered double SE TOCSY-TOCSY spectrum taken at 2.39 ppm corresponding to the $H_{\gamma 2}$ of E25 and the $H_{\gamma 2/\gamma 3}$ of E20 (b), the successive plane at 2.35 ppm corresponding to the $H_{\gamma 2}$ of E23 (not visible here) (c) and at 1.93 ppm corresponding to the $H_{\beta 2/\beta 3}$ of E27 (d). Again, the vertical dashed lines indicate the H_α frequencies. The increased resolution of the 3D spectrum allows for the complete assignment of these four glutamic acids. Peaks on the single-transfer-line (Vuister et. al., 1990) are labeled with a cross; genuine 3D-peaks are labeled with a filled circle. The labels indicate the assignment of the cross peaks in the F2-dimension. The peak labeled with a star corresponds to the transfer $H_{\beta 2} \rightarrow H_{\beta 2} \rightarrow H_\alpha$ of R24. The peak labeled with a hash is an artifact. This is easily recognized from its antiphase line shape in F3 and F1 (not shown). All cross peaks are non-stereospecifically assigned. Only positive contour levels are plotted.

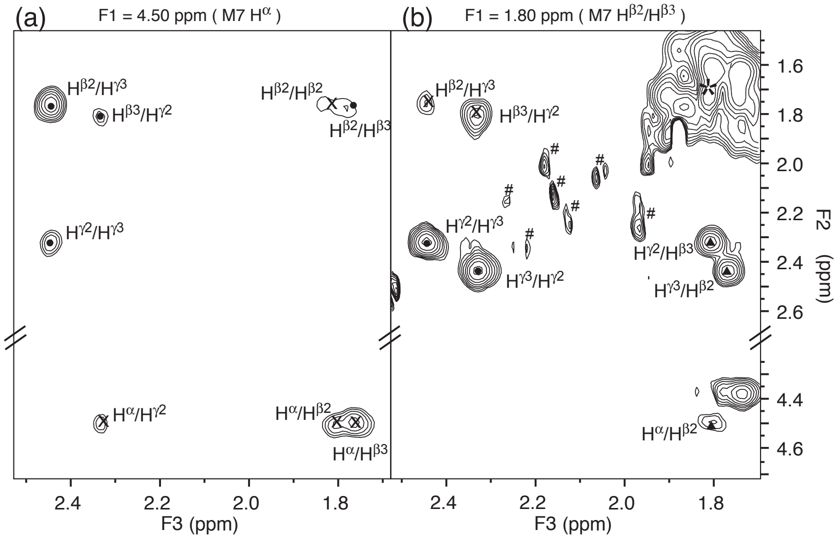


Figure 3.3. Sections of (F2,F3)-planes of the 3D $^{13}C/^{15}N$-filtered double SE TOCSY-TOCSY spectrum taken at 4.50 ppm corresponding to the H_α frequency of M7 (a) and at 1.80 ppm corresponding to the $H_{\beta2/\beta3}$ frequency of M7 (b). Again, peaks on the single-transfer-line are labeled with a cross, genuine 3D-peaks are labeled with a filled circle and back-transfer peaks (Vuister et. al., 1990) are labeled with a filled triangle. The star indicates the intersection of the 3D body diagonal with the plane. While the region between 1.8 and 2.5 ppm is very crowded in the 2D X-filtered TOCSY spectrum, this is not true for these planes, allowing an easy and complete identification of this spinsystem. The labels indicate the assignment of the cross peaks in the F2- and F3-dimension. The peaks labeled with a hash result from side lobes in the F1-dimension of intense diagonal peaks.

spectrum taken at 2.39, 2.35 and 1.93 ppm, corresponding to the $H_{\gamma2}$ E25/$H_{\gamma2/\gamma3}$ E20, $H_{\gamma2}$ E23 and the $H_{\beta2/\beta3}$ E27, respectively. From these sections, it becomes apparent that four spin systems give rise to the cross peaks, of which the H_α frequencies are indicated by the dashed vertical lines. For instance, the upper peak at (F1, F2, F3 = 2.39, 2.01, 4.10 ppm) in panel (b) corresponds to the transfer $H_{\gamma2} \rightarrow H_{\beta2/\beta3} \rightarrow H_\alpha$ of E25. The $H_{\gamma2/\gamma3}$ of E20 also resonates in this plane of the 3D spectrum, giving rise to a peak corresponding to the transfer $H_{\gamma2/\gamma3} \rightarrow H_{\beta2/\beta3} \rightarrow H_\alpha$ and the single transfer peak $H_{\gamma2/\gamma3} \rightarrow H_{\gamma2/\gamma3} \rightarrow H_\alpha$. Likewise, the peaks in panel (c) correspond to the transfer $H_{\gamma2} \rightarrow H_{\beta2/\beta3} \rightarrow H_\alpha$ and $H_{\gamma2} \rightarrow H_{\gamma3} \rightarrow H_\alpha$ of E23 and the lower peak in panel (d) corresponds to the transfer $H_{\beta2/\beta3} \rightarrow H_{\gamma2} \rightarrow H_\alpha$ of E27.

Another example of the increased resolution of the 3D X-filtered TOCSY-TOCSY is the assignment of M7. The methylene protons of the five glutamic acids, the glutamine, and the two methionines of the extended SID-fragment all resonate in same region from roughly 1.8 to 2.5 ppm. Consequently, cross peaks between the methylene protons of M7 were obscured in this very crowded region in the 2D X-filtered TOCSY spectrum. This made assignment of M7 impossible, as cross peaks between the H_α and the methylene protons could not be identified as well. Adding a third dimension, however, the cross peaks between the methylene protons can also be generated starting at the H_α of M7. This is shown in panel (a) of Figure 3.3, in which peaks corresponding to the transfer $H_\alpha \rightarrow H_{\beta2} \rightarrow H_{\gamma3}$ and $H_\alpha \rightarrow H_{\beta3} \rightarrow H_{\gamma2}$ can easily be identified. In a similar fashion, cross peaks between $H_{\gamma2}$ and $H_{\gamma3}$ can easily be identified by starting from the $H_{\beta2}$ and $H_{\beta3}$, as is illustrated in panel (b) of Figure

3.3. These cross peaks are unidentifiable in a 2D spectrum, since they are very close to the crowded diagonal.

Using the above approach and a 2D $^{13}C/^{15}N$ double filtered NOESY spectrum (not shown) we obtained the complete sequential assignment of the SID24 peptide bound to the PAH2 domain, which is an essential step in solving the structure of the complex between the two biomolecules.

Conclusion

In conclusion, our 3D double SE X-filtered TOCSY-TOCSY is an effective tool for the assignment of unlabeled molecules with an unusual size or composition. Spectra can be recorded in an acceptable amount of time on millimolar samples of complexes of reasonable size. The third dimension and the inclusion of double sensitivity enhancement extends the limits of the regular X-filtered TOCSY experiment to allow for resonance assignment in cases where extensive overlap prevents the use of conventional two-dimensional methods.

Data bank accession numbers

The assignments of the extended 24-residue SID-peptide are deposited at the BioMagRes bank under entry number BRMB-6102.

Chapter 4.

Hugo van Ingen,
Edwin Lasonder,
Jacobus F.A. Jansen,
Anita M. Kaan,
Christian A.E.M. Spronk,
Henk G. Stunnenberg
and Geerten W. Vuister

Biochemistry
(2004), **43**, p. 46-54

Extension of the binding motif of the Sin3 interacting domain of the Mad-family proteins

Acknowledgements. We thank Jan Aelen for technical assistance and purification of the labeled proteins and Tine Walma for proofreading the manuscript. This research was funded by a “Jonge Chemici” grant by the Netherlands Organization for Scientific Research, NWO-CW (JC 99-03) to G.W.V.

Sin3 forms the scaffold for a multi-protein co-repressor complex that silences transcription via the action of histone deacetylases. Sin3 is recruited to the DNA by several DNA binding repressors, such as the helix-loop-helix proteins of the Mad family. Here, we elaborate on the Mad-Sin3 interaction based on a binding study, solution structure and dynamics of the PAH2 domain of mSin3 in complex to an extended Sin3 Interacting Domain (SID) of 24 residues of Mad1. We show that SID-residues Met7 and Glu23, outside the previously defined minimal binding motif, mediate additional hydrophobic and electrostatic interactions with PAH2. On the basis of these results we propose an extended consensus sequence describing the PAH2-SID interaction specifically for the Mad family, showing that residues outside the hydrophobic core of the SID interact with PAH2 and modulate binding affinity to appropriate levels.

Introduction

Sin3 is a ubiquitous, highly conserved protein involved in repression of gene transcription. Sin3 acts as a co-repressor and is recruited to the DNA by various DNA binding repressors. Gene silencing is brought about by its association with histone deacetylases (HDAC) (Hassig et al., 1997; Laherty et al., 1997; Zhang et al., 1997), which deacetylate the core histones resulting in a repressed state of the chromatin (Ayer, 1999; Ahringer, 2000). Sin3 is thought to act as a scaffold protein for the co-repressor complex, recruiting not only HDAC1/HDAC2, but also histone binding proteins RbAp46/RbAp48, Sin3 associated proteins SAP18/SAP30 and the recently identified Sds3, SAP180 and SAP130 to form a large multimeric active repression complex (Hassig et al., 1997; Laherty et al., 1997; Zhang et al., 1997; Zhang et al., 1998; Alland et al., 2002; Fleisher et al., 2003).

Sin3 contains at least three stretches of about 80 residues, subsequently denoted as PAH1, PAH2 and PAH3 (poly amphipathic helix domains), which are highly conserved in different forms of the Sin3 protein. These domains are suggested to be the protein-protein interaction domains by which Sin3 fulfils its role as a scaffold. However, thus far the PAH2 domain has mainly been identified as the interaction domain, whereas interactions involving the PAH1 and PAH3 domains are much more tentative. Moreover, for a number of interactors, regions other than the PAH domains were shown to be essential for binding as in the case of HDAC (Laherty et al., 1997), Sds3 (Alland et al., 2002), p53 (Murphy et al., 1999) and MeCP2 (Nan et al., 1998; Jones et al., 1998).

A large number of repressors have been identified that by direct interaction with Sin3 recruit the co-repressor complex; one of which is Mad1 (Ayer et al., 1995, Schreiber-Agus et al., 1995), a basic region/ helix-loop-helix/leucine zipper (bHLHZip) transcription factor. Mad1 belongs to a family of four proteins (Mad1, Mxi1, Mad3 and Mad4), which are part of the Myc/Max/ Mad network involved in cell cycle regulation. Both the Mad proteins and the proto-oncogene product Myc bind E-box DNA sequences (CACGTG) and heterodimerize with Max through their bHLHZip domains. While Myc-Max promotes cell proliferation and transformation and is implicated in tumorigenesis, Mad-Max is found at high levels in differentiating cells and represses a subset of target genes of Myc (Zhou and Hurlin, 2001; Lüscher, 2001; Grandori et al., 2000). Abrogation of Mad-mediated repression has been linked to the development of promyelocytic leukemia (Kahn et al., 2001) and knock-out studies implicated Mxi1 to be a potential tumor suppressor (Foley and Eisenman, 1999).

The repressive function of the Mad family depends on their interaction with the PAH2 domain of Sin3 through a short amphipathic helix located at their N-terminus termed SID (Schreiber-Agus and DePinho, 1998; Eilers et al., 1999) for Sin3 Interacting Domain. Recently, a molecular basis has been provided for understanding this interaction by the NMR structure of the Mad1-SID Sin3-PAH2 complex (Brubaker et al., 2000; Spronk et al., 2000). In the study of Brubaker et al. (2000) the PAH2 domain of the A-form of the mammalian Sin3, mSin3A, was used together with a SID of 16 residues (residues 6-21), which will subsequently be denoted as PAH2^{A}-SID16. The study of Spronk et al. (2000) used the PAH2 domain of the B-form of mammalian Sin3, mSin3B, and a SID of 13 residues (residues 8-20), subsequently denoted as PAH2^{B}-SID13. The PAH-domains of mSin3A and mSin3B share ~60-70% sequence identity and at present, there is little knowledge about any functional differences between these two forms of mSin3. In both studies, the PAH2 domain is folded as a four-helix bundle and the amphipathic helix of the SID is wedged between helix α1 and helix α2 of PAH2 at angle of ~45°, a folding motif denoted as the "wedged helical bundle" (Spronk et al., 2000).

On the basis of their structure, multiple sequence alignments and mutation data, Brubaker et al. (2000) proposed a consensus sequence for a 13-residue minimal SID, showing that interaction is mainly mediated by the hydrophobic surface of the amphipathic helix. However, several binding assays including our own (*vide infra*), indicate that extension of the Mad1-SID increases the affinity for PAH2 (Eilers et al., 1999; Brubaker et al., 2000), possibly by additional intermolecular interaction mediated by conserved residues outside the minimal binding motif.

Here, we elaborate further on the structural details of the Mad-Sin3 interaction by reporting the results of a binding study, a high-resolution structure and a dynamics study on the complex between an extended SID of 24 residues of Mad1 and the PAH2 domain of mSin3B, subsequently denoted as PAH2^{B}-SID24. We show that conserved SID-residues outside the minimal binding motif, *viz.* Met7 and Glu23, interact with PAH2 and thus contribute to the binding affinity, specifically for the Mad family.

Results

Binding

To investigate the effect of extension of the SID on binding affinity, we performed SPR measurements on four different Mad1-SID constructs extended at the C-terminus (Mad1-SID5-20, 5-24, 5-28 and 5-35). The results are listed in Table 4.1 and show that extension with the charged residues 21-24 (RRER) increases affinity approximately 5-

Table 4.1.
Binding affinity of Mad1-SID for the PAH2 domain of mSin3B

Mad1-SID	sequence	K_D (μM)
5-20	VRMNIQMLLEAADYLE	1.4 ± 0.6
5-24	VRMNIQMLLEAADYLERRER	0.4 ± 0.1
5-28[a]	VRMNIQMLLEAADYLERREREAEH	0.3 ± 0.2
5-35	VRMNIQMLLEAADYLERREREAEHGYASML	0.2 ± 0.1

[a] The Mad1-SID construct used for the PAH2^{B}-SID24 complex

fold, suggesting an electrostatic interaction between this part of the SID and PAH2. Residues 25-35 do not seem to significantly contribute to binding affinity as constructs SID5-28 and SID5-35 have similar binding affinities compared to SID5-24.

Using ITC measurements Brubaker et al. (2000) found a K_D of ~0.029 μM for SID6-21 and ~0.015 μM for SID1-35 binding the PAH2 domain of mSin3A, values that are significantly lower when compared to those of mSin3B. We speculate that these differences are most likely caused by the different methods used, since the structure of PAH2^{A}-SID16 shows only little differences with the PAH2^{B}-SID13 and PAH2^{B}-SID24 structures (*vide infra*). Unfortunately, Brubaker et al. do not provide experimental details regarding the ITC measurements. Nevertheless, both studies clearly show that despite the relative differences in the values of the K_D an extension of SID yields a higher affinity.

Structure

The structure of the complex between the extended 24-residue SID and PAH2 was determined using high-resolution multi-dimensional heteronuclear NMR. Analysis of edited- and filtered-NOESY experiments revealed 11 new intermolecular NOEs between PAH2 and residues in the extension part of the SID, such as between the ε-methyl

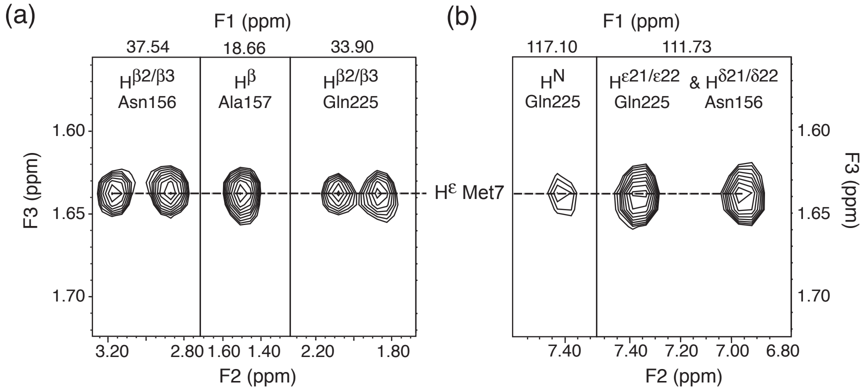


Figure 4.1. New intermolecular NOEs in the extended PAH2^{B}-SID24 complex. (a) selected strips from the 3D (^{13}C-edited-^{13}C/^{15}N-filtered)-HMQC-NOESY spectrum and (a) selected strips from the 3D (^{15}N-edited-^{15}N/^{13}C-filtered)-HSQC-NOESY spectrum. Frequencies in the F1-dimension and the assignment of the peaks in the F2-dimension are given at the top of each strip. Shown in (a) are the peaks corresponding to the NOE transfer from the side chain β-protons of PAH2 residues Asn156, Ala157 and Gln225 to the ε-methyl group of Met7 in the SID. Shown in (b) are the peaks corresponding to the NOE transfer from the backbone amide-proton of Gln225 and the side chain amide-protons of residues Asn156 and Gln225 in PAH2 to the ε-methyl group of Met7 in the SID.

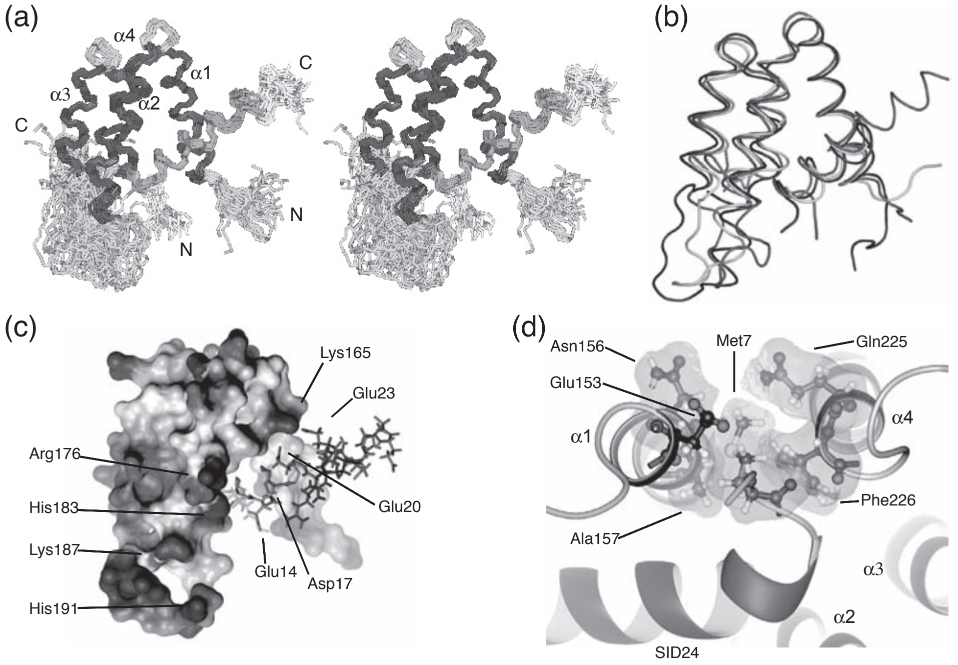


Figure 4.2. (a) stereo diagram of the structural ensemble of 30 structures of the PAH2^B-SID24 complex. Helices of PAH2^B are shown in black, the helix of the SID is shown in dark gray; coil and turn regions are shown in light gray for PAH2^B and the SID. (b) structural superposition of the lowest energy structures of the PAH2^B-SID13 (light gray; PDB code 1E91), PAH2^B-SID24 (dark gray; PDB code 1PD7) and PAH2^A-SID16 (black; PDB code 1G1E) complexes. All structures were superposed on the PAH2 domain. (c) additional electrostatic interactions between PAH2^B-Lys165 and SID-Glu23. PAH2^B is shown in surface representation with color coding according to electrostatic potential; the SID is shown in stick-representation with its charged residues shown (for both molecules with color coding: black = positively charged, dark gray = negatively charged). (d) additional interactions between SID-Met7 and residues in α1 and α4. The contact-surface of these residues is shown in light gray. This figure was prepared using MOLMOL (Koradi et al., 1996).

group of Met7 and the side chain β-protons of Asn156 and Gln225 (cf. Figure 4.1). In total 1762 unique distance restraints (Table 4.2) and 140 dihedral angle restraints for the backbone angles φ, ψ were used to calculate the solution structure of this complex. A stereo view of the ensemble of structures is presented in Figure 4.2(a). Statistical data for the ensemble are reported in Table 4.2. All structures contain neither distance violations > 0.5 Å nor dihedral angle violations > 5° from experimental data. The pairwise heavy backbone atom RMSD is 0.8 Å for the ordered parts of the structure (residues 152-189 and 202-227 for PAH2 and residues 7-24 for the SID), while the pairwise all heavy atom RMSD is 1.7 Å for the same residues, indicating a well-defined ensemble. The Ramachandran plot displays 84% of the residues in most favored regions and 15% in additionally allowed regions.

The complex forms a wedged helical bundle in which the amphipathic helix of the SID is located between helices α1 and α2 of the four-helix bundle of the PAH2 domain at an angle of ~45° with respect to α2. Residues 152-167, 172-189, 202-212 and 217-226 of PAH2 form the four helices α1, α2, α3, and α4 of the four-helix bundle. Residues 168-171 and 213-216 fold into turns and residues

Table 4.2.
Structural statistics for the PAH2^{B}-SID24 complex

A. Restraint information	
total no. of distance restraints	1762
intra-PAH2 (intra-residual/sequential/medium/long)	1324 (557/347/296/123)
intra-SID (intra-residual/sequential/medium/long)	312 (155/83/74/0)
intermolecular	126
dihedral angle restraints	140
B. Average RMS deviation from experimental restraints	
all distance restraints (Å)	0.056 ± 0.002
all dihedral angle restraints (deg)	0.58 ± 0.01
C. Pairwise Cartesian RMS deviation (Å)	
ordered backbone heavy atoms	0.77 ± 0.12
ordered all heavy atoms	1.74 ± 0.17
global backbone heavy atoms	2.87 ± 0.61
global all heavy atoms	3.73 ± 0.56
D. Ramachandran quality parameters (%)	
residues in most favored regions	84.3
residues in allowed regions	12.1
residues in additionally allowed regions	2.6
residues in disallowed regions	1.1
E. Abnormalities found in structural checks	
abnormally short interatomic distances	3
unsatisfied H-bond acceptors (buried)	4
unsatisfied H-bond donors (buried)	12
F. Average RMS deviation from current reliable structures (RMS Z-scores, null deviation = 1)	
bond lengths	1.013
bond angles	1.296
omega angle restraints	1.451
side chain planarity	1.113
improper dihedral distribution	1.007
inside/outside distribution	1.001
G. Average deviation from current reliable structures (Z-scores, null deviation = 0)	
second generation packing quality	-2.157
Ramachandran plot appearance	-3.354
chi-1/chi-2 rotamer normality	-2.092
backbone conformation	-1.658

190-201 form a large loop between helices α2 and α3. Residues 233-252 were not included in the structure calculation as NMR data for this highly flexible region (*vide infra*) did not give any indication of secondary structure. Residues 9-24 of the SID fold as an α-helix, with its hydrophobic residues intensively contacting the hydrophobic pocket of PAH2.

To compare the structure of PAH2^{B}-SID24 with our previous structure of PAH2^{B}-SID13, a structural superposition was made using the lowest energy structures (Figure 4.2(b)). The heavy backbone atom RMSD for the ordered regions is 1.2 Å and the heavy all-atom RMSD is 2.0 Å for the same regions, similar to the values for the ensemble of PAH2^B-SID24 structures. As these two structures are highly similar we refer to Spronk et al. (2000) for a detailed description of the fold and the hydrophobic interactions stabilizing the complex. Here, we will focus on the structural role of residues located outside the previously defined minimal binding motif.

At the C-terminal end of the Mad minimal binding motif, there are eight additional residues: RREREAEH (residues 21-28). Charged residues 21-24 form a solvent exposed extension of the SID-helix comprising almost two helical turns. These residues were identified in our binding study as responsible for a 5-fold increase in binding affinity. Our structure shows that this increase is due to the electrostatic interaction between the intermolecular ion pair formed by SID-Glu23 and PAH2-Lys165 (Figure 4.2(c)). Table 4.3 summarizes an analysis of this ion pair in the ensemble of conformers obtained by discriminating between a hydrogen bond interaction, a N-O bridge and a long range electrostatic interaction (Kumar and Nussinov, 2002). The results show that in 60% of the conformers Lys165 is hydrogen bonded to Glu23. Furthermore, Lys165 also forms an intermolecular ion pair with SID-Glu20, although this interaction is mainly long range. Figure 4.2(c) displays a surface representation of the PAH2 domain color coded according to the electrostatic potential, showing that helix α2 and part of the loop α2-to-α3 are mainly positively charged due to the presence of Arg176, His183, Lys187 and His191. The SID contains a negatively charged rim formed by residues Glu14, Asp17 and Glu20, resulting in a network of long range electrostatic interactions between these residues. Other charged residues in the C-terminal extension, residues Arg21 and Arg22 and residues Arg24 to His28 point away from the interaction surface, into the solvent.

Table 4.3
Analysis of intermolecular interactions involving PAH2-Lys165.

ion pair	interaction type[a]	N (%)[b]
PAH2-Lys165 – SID-Glu23	hydrogen bonded N-O bridge long range ion pair	18 (60%) 4 (13%) 8 (27%)
PAH2-Lys165 – SID-Glu20	hydrogen bonded N-O bridge long range ion pair	3 (10%) 0 (0%) 27 (90%)

[a] Interaction in an ion pair is classified into hydrogen bonded, N-O bridges or long-range ion pairs, depending on their geometry. Criteria for hydrogen bonds (D-H···A) were a maximum distance r(H,A) of 2.5Å and a maximum angle θ(D-H-A) of 90°. Interaction is defined as an N-O bridge when r(N,O) < 4.0Å, otherwise it is classified as a long range ion pair (Kumar and Nussinov, 2002).

[b] The number of conformers in which the specified interaction type is observed. The percentage of all conformers in the ensemble is shown in parentheses.

The three additional residues at the N-terminus, residues 5-7 (VRM), do not form an extension of the SID helix; instead the backbone adopts a mainly extended conformation protruding from the base of PAH2. Consistent with the intermolecular NOEs observed for SID-Met7, its side chain is inserted between the backbone of α1 and α4 in a hydrophobic pocket formed by Glu153, Asn156 and Ala157 in helix α1 and Gln225 and Phe226 in helix α4 (Figure 4.2(d)).

Another additional interaction was found for residue SID-Met11, which is in the minimal binding motif. New intermolecular NOEs were observed indicating that the side chain of SID-Met11 contacts residues Glu153, Phe154 and Ala157 in helix α1 (data not shown). Comparison of the observed intermolecular NOEs in the PAH2^B-SID24 and PAH2^B-SID13 complexes yielded no further significant differences.

Dynamics

We performed ^{15}N relaxation measurements to probe the dynamical behavior of the PAH2 domain when complexed to the extended SID of 24 residues and the minimal SID of 13 residues. We recorded the ^{15}N-T_1, ^{15}N-$T_{1\rho}$ and $\{^1H\}$-^{15}N NOE experiments and used the model-free approach (Lipari and Szabo, 1982ab; Clore et al., 1990) as implemented in the program Modelfree 4.15 (Mandel et al., 1995; Palmer et al., 1991) to calculate dynamical parameters describing the global rotational diffusion of the complex and local internal motions of the backbone of PAH2.

For the PAH2^B-SID24 complex, reliable data could be extracted for 77 out of 102 non-proline residues; for the PAH2^B-SID13 complex we obtained data for 69 out of 102 residues. The global rotational diffusion tensor was determined to be isotropic with optimized values for the global correlation time, τ_c, of 11.5 ± 0.1 and 10.0 ± 0.1 ns for the PAH2^B-SID24 and the PAH2^B-SID13 complex, respectively. These values are higher than expected for a typical protein of similar size, but just within the upper limit (Engelke and Rüterjans, 1999). A possible explanation for this deviation is a reduction of the rotational diffusion caused by the unstructured flexible C-terminal tail and the large flexible loop between α2 and α3 in PAH2 indicated by low values for the NOE (average 0.35 ± 0.09 for the loop and –0.6 ± 0.2 for residues 233-252) and the order parameter (average 0.55 ± 0.18 for the loop and 0.05 ± 0.02 for residues 233-252).

Model-free analysis describes local motions using three dynamical parameters: the square of the generalized order parameter S^2, which is correlated to the amplitude of the motion, the effective correlation time of the internal motions τ_e, which indicates the time scale of internal motions and the chemical exchange contribution R_{ex}, which is indicative of motions on a microsecond to millisecond time scale. These parameters are presented in Figure 4.3 for both the PAH2^B-SID24 and the PAH2^B-SID13 complex. The four helices and the two turns α1-to-α2 and α3-to-α4 have average S^2-values of around 0.9, indicating that these regions form rigid structures. The loop between α2 and α3 and the N- and C-termini are the highly flexible parts of the domain with average S^2 values

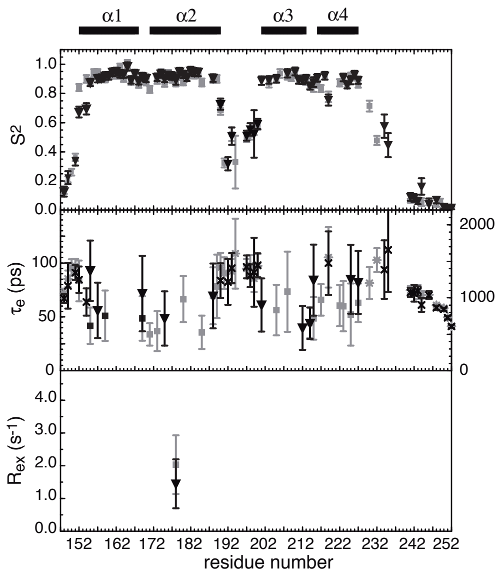


Figure 4.3. Results of the ^{15}N relaxation analysis for the PAH2 domain using the model-free approach. Shown are the dynamical parameters of PAH2 in the PAH2^B-SID24 complex (grey squares) and in the PAH2^B-SID13 complex (black triangles). The top panel displays the generalized order parameter squared (S^2), the middle panel shows the rotational correlation times (τ_e) for fast internal motions (grey squares for PAH2^B-SID24 and black triangles for PAH2^B-SID13; axis shown on the left-hand side) and for slow internal motions (grey stars for PAH2^B-SID24 and black crosses for PAH2^B-SID13; axis shown on the right-hand side) and the lower panel presents the chemical exchange contribution R_{ex}. The secondary structure of the PAH2^B-domain is shown in the top of the figure with helices represented as black rectangles.

in the range of 0.2-0.5, consistent with the higher local RMSD for these regions in the ensemble.

Nineteen residues for PAH2^{B}-SID24 and 11 residues for PAH2^{B}-SID13 show contributions from local internal motions on a time scale of 30-150 ps. Mostly, these residues have surface exposed side chains. One residue, Phe178 in α2, requires a small chemical exchange contribution in both complexes (2.0 ± 0.9 s^{-1} for PAH2^{B}-SID24 and 1.4 ± 0.8 s^{-1} for PAH2^{B}-SID13). Strikingly, Leu219 in helix α4 shows a complex dynamical behavior with a lower than average S^2 (~0.76) and internal motions on a fast and slow (~1.5 ns) time scale. Furthermore, residues Lys163 in the PAH2^{B}-SID24 complex and Lys165 in both complexes in helix α1 display a significantly higher than average S^2.

With the exception of the N-terminal region of helix α1, the dynamics of PAH2 in the PAH2^{B}-SID13 and PAH2^{B}-SID24 complexes is highly similar, as expected since the fold of PAH2 is identical. Dramatic changes are observed for residues 152 and 154 in the first turn of helix α1, which show a large increase in S^2 of about 0.2 units for the PAH2^{B}-SID24 compared to PAH2^{B}-SID13. While in the PAH2^{B}-SID13 complex these residues are flexible and have internal motions on a fast and a slow time scale, in the PAH2^{B}-SID24 complex they are significantly less flexible and do not show internal motions on a slow time scale. This is in close agreement with additional intermolecular interactions observed in the PAH2^{B}-SID24 complex between Met7 and Met11 with the N-terminal region of α1, which will evidently constrain the flexibility of the backbone. These data, together with the NOE relaxation data of the free PAH2 domain (Spronk et al., 2000) suggest that the flexibility of the first two turns of α1 in the free PAH2 is necessary to allow the SID to enter the binding pocket. Upon binding, intermolecular interactions between the SID and the N-terminal part of helix α1 'lock' the helix, increasing its rigidity and stability.

Chemical shift perturbation

Figure 4.4 displays the chemical shift differences between the PAH2^{B}-SID24 and PAH2^{B}-SID13 complexes for the backbone resonances (Figure 4.4(a)) and the side chain amide resonances of asparagines and glutamines (Figure 4.4(b)). Consistent with the structural differences between the two complexes, it shows large differences for the first part of helix α1 and the last few residues of α4, with respect to the backbone resonances and side chain resonances. Especially the backbone resonances of Glu153 and Gln225 and the side chain amides of Asn156 and Gln225 shift dramatically, which is in agreement with the observed new intermolecular contacts between SID-Met7 and these resi-

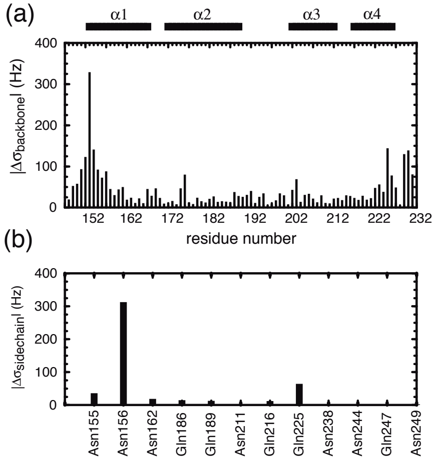


Figure 4.4. Absolute chemical shift differences between the PAH2^{B}-SID24 and PAH2^{B}-SID13 complexes for backbone resonances (a) and selected side chain resonances (b). The difference in chemical shift was calculated as the vector displacement in hertz between the two complexes of the HN, N and C$_{\alpha}$-resonances (a) or the C$_{\gamma}$, N$_{\delta}$, H$_{\delta21}$ and H$_{\delta22}$ for the asparagine amide group and the C$_{\delta}$, N$_{\varepsilon}$, H$_{\varepsilon21}$ and H$_{\varepsilon22}$ for the amide group of glutamines (b). The secondary structure of the PAH2^{B}-domain is shown in the top of the figure with helices represented as black rectangles.

dues. Residues 229-232 also show significant differences in backbone chemical shift, but NOE data did not give any indication of a structural change for these residues.

Discussion

SID-Met7 and SID-Glu23 contribute to binding affinity

We reported a binding study, the solution structure and dynamics of the complex between a 24-residue extended Sin Interacting Domain of Mad1 and the PAH2 domain of mSin3B. The $PAH2^B$-SID24 structure shows that extension of the SID does not affect the global fold of PAH2 and preserves the general mode of the SID-PAH2 interaction. However, new intermolecular interactions, both hydrophobic and electrostatic, involving residues in the N- and C-terminal parts of the SID have been identified, i.e. SID-Met7 and SID-Glu23. On the basis of the observed intermolecular NOEs and in agreement with chemical shift perturbation data, SID-Met7 is involved in a hydrophobic interaction with residues in α1 and α4, resulting in an increase in rigidity of the first turn of α1 in accordance with our dynamics data. Furthermore, there is a new electrostatic interaction and intermolecular hydrogen bond involving SID-Glu23 and PAH2-Lys165 in α1 of PAH2, as suggested by both our SPR binding assay and the structural ensemble. Electrostatic interactions are notoriously difficult to prove directly by NMR data. Also in this case, we did not observe intermolecular NOEs or significant changes in chemical shifts involving PAH2-Lys165. However, the use of a water refinement protocol that explicitly treats electrostatic forces will correct for this lack in experimental restraints. In fact, the CHARM22 protocol has been proven to yield more accurate and precise NMR structures with both significantly improved electrostatic and hydrogen-bond donor and acceptor properties (Spronk et al., 2002).

Apart from the crucial role for PAH2-Lys165 in mediating intermolecular contacts, PAH2 residues Glu153, Asn156, Ala157, Gln225 and Phe226 also contribute to the additional intermolecular interactions.

Comparison of $PAH2^B$-SID24 with $PAH2^A$-SID16

Figure 4.2(b) shows the superposition of the $PAH2^B$-SID24 structure and the $PAH2^A$-SID16 structure (Brubaker et al., 2000). The $PAH2^B$ and $PAH2^A$ sequences have an overall sequence identity of 61%, which increases to 75% for the structured regions (residues 152-189 and 202-226). The α2-α3 loop shows most sequence diversity and is three residues longer in $PAH2^A$ compared to $PAH2^B$. Overall the fold is highly similar. There are, however, three striking differences: *i*) helix α2 is extended by almost two turns in $PAH2^A$-SID16; *ii*) SID-Met7 does not interact with α1 and α4 in $PAH2^A$-SID16; *iii*) SID-Glu20 is identified as the key interaction partner for Lys315 in $PAH2^A$-SID16, corresponding to Lys165 in $PAH2^B$.

While helix α2 in $PAH2^B$-SID24 is 18 residues long and extends from Pro172 to Gln189, this helix is extended at its C-terminus in $PAH2^A$-SID16 to cover 24 residues. Comparison of the corresponding sequence of mSin3B, LHTKGR (residues 190-195), to this C-terminal extension of helix α2 in mSin3A, RNAKEA (residues 340-345), suggests an intrinsic difference as the mSin3A sequence has a higher helix-propensity than the mSin3B sequence. Furthermore, our dynamics data clearly indicate that α2 extends to residue 189 and that residues 190 to 200 form a highly flexible loop. Brubaker et al. (2000) do not provide detailed information on the dynamics of the $PAH2^A$ domain, but do state that the C-terminal part of helix α2 has lower NOE values (in the range of 0.3-0.7) indicating increased flexibility.

Interaction between SID-Met7 and α1 and α4 of PAH2 is not present in the $PAH2^A$-SID16 structure. Analysis of the restraint-lists of this structure shows that Brubaker et al. did not identify any intermolecular NOEs for Met7. Possibly, Met7 is too flexible to interact, since it is the penultimate residue in the mSin3A complex. Nevertheless, our studies show convincing evidence for this interaction: *i*) observation of a number of intermolecular NOEs between Met7 and residues in α1 and α4; *ii*) increase in rigidity of the first part of α1 as a result of this interaction; *iii*) large chemical shift changes for the involved residues. Furthermore, biochemical data as reported by Eilers et al. (1999) confirms this interaction. In their two-hybrid assay, extension of the minimal SID (residues 8-20) with Met7 yielded a 1.7-fold increase in reporter activity, clearly suggesting a higher binding affinity. Finally, sequence alignment of the Mad family members in Figure 4.5(a) shows that a large hydrophobic residue is conserved at position 7. Furthermore, a small residue with a preference for turns is conserved at position 8, which is crucial to allow proper positioning of Met7.

Previously, Brubaker et al. (2000) identified the interaction between PAH2^A-Lys315 and SID-Glu20 as a key intermolecular interaction. Mutation of PAH2^A-Lys315 to either Ala or Glu resulted in a significantly reduced binding affinity as measured in a GST pull-down assay using a GST-fused SID as bait for either wild type or mutant mSin3A. However, a SID consisting of the first 27 residues of Mad1 was used, thus including SID-Glu23. Our structure clearly shows that SID-Glu23 is much closer to, and thus more strongly interacting with PAH2^B-Lys165 than SID-Glu20. Both in our structure and in their PAH2^A-SID16 structure, SID-Glu20 forms mainly a weak, long range ion pair with this lysine residue. Therefore, we suggest that SID-Glu23 and not SID-Glu20 is responsible for this key intermolecular interaction.

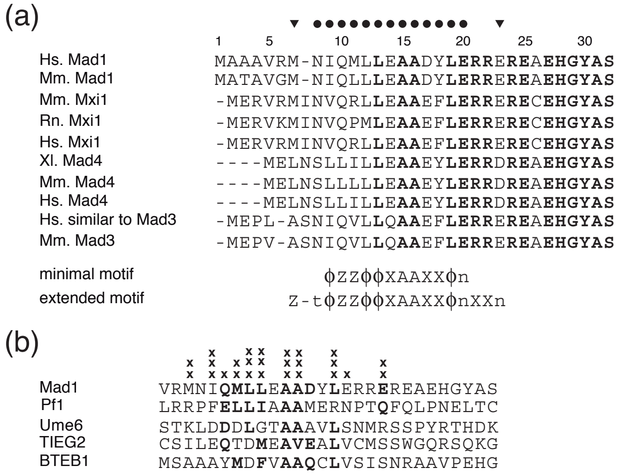


Figure 4.5. (a) multiple sequence alignment of the Mad family proteins from different species. Numbering in top row refers to *Hs.* Mad1. Conserved residues are depicted in bold type. ● = residues in the minimal binding motif; ▼ = residues showing additional intermolecular interactions. The minimal and extended binding motif are given below the sequence alignment (written in standard one-letter code and ϕ = bulky hydrophobic residue; X = non-proline residue; Z = large residue with a significant aliphatic component; n = negatively charged residue; t = small residue). (b) alignment of sequences of the PAH2 binding SIDs of Mad1, Pf1, Ume6, TIEG2 and BTEB1. Sequence similarities are depicted in bold type. The number of crosses above a residue indicates its relative importance in mediating intermolecular contacts with the PAH2 domain for the case of Mad1: XXX = most important, XX = important, and X = less important.

Extended binding motif

The impact of these interactions on binding affinity when compared to the minimal binding sequence can be estimated to be an increase of approximately 10-fold based on our interaction study and that of Eilers et al. (1999), in which the former shows a 5-fold increase caused by the interactions involving Glu23 and the latter suggests that Met7 is responsible for an almost 2-fold increase in binding affinity. Therefore, we propose to extend the sequence motif to describe the Mad-Sin3 interaction more accurately. The sequence alignment of the Mad family proteins shown in Figure 4.5(a) shows that a negative charge is conserved at the position of Glu23 and a bulky aliphatic residue is conserved at the position of Met7 together with a small turn-residue at the position of Asn8, which is crucial because of its structural role in allowing the SID backbone to bend in the direction of α1 and α4. For Mad4 proteins there are two small turn-residues, Asn and Ser, that could allow Leu to take the role of Mad1-Met7 in mediating hydrophobic contacts with α1 and α4. In the Mxi1 proteins there is an Ile positioned between the Asn and Met, suggesting that in this case Ile rather than Met is involved in contacts with the protein.

This high degree of conservation suggests that the binding mode described here for the Mad1-Sin3 interaction is also valid for the other Mad family proteins. The extended sequence motif is ZtϕϕZZϕϕXAAXXϕnXXn, where ϕ denotes any bulky hydrophobic residue, X is any non-proline residue, Z denotes any hydrophobic or polar/charged residue with a significant aliphatic component, t represent a small turn residue and n designates a negatively charged residue.

The resulting view is that the hydrophobic face of the SID formed by Ile9, Leu12, Leu13, Ala15, Ala16 and Leu19 provides the basis for interaction with the hydrophobic pocket of PAH2, resulting in a basic level of binding affinity. Other residues, outside this minimal hydrophobic surface, *viz.* Met7, Met11, Glu23 and to a lesser extent Glu20 and Gln10, provide an additional contribution to affinity. The remaining residues are excluded from the sequence motif as they show only minor involvement in intermolecular contacts. Residues Glu14 and Asp17 participate only in long-range electrostatic interactions, which might not stabilize the complex, as long-range electrostatic interactions are weak and the presence of a water shell could quench the effective electrostatic field even further. However, since a network of electrostatic interactions is possible, involving the four negatively charged residues in the SID and at least four positively charged residues in PAH2, the net-effect can be substantial. Additionally, these charges could play a role in complex formation, enhancing the on-rate by long-range electrostatic forces.

The functional role of the 100% conserved EHGYAS sequence is still unclear. One might envisage that the GYASMLP sequence extends the Mad helix described here for another two turns. This will leave the EGYAS sequence too far away from PAH2 for any direct interaction. This is confirmed in the SPR binding assay, in which extension of the SID with this sequence did not affect the binding affinity significantly, in agreement with earlier reports (Ayer et al., 1995, Eilers et al., 1999). Thus, we speculate that it plays an important architectural role in the complete Mad protein or might interact with an unidentified partner.

Implications for other PAH2-interacting SIDs

The extended sequence motif not only is a more accurate description of the interaction of proteins from the Mad family with PAH2, but also is specific for this family. Although other repressor proteins also have PAH2 interacting SID sequences that can be modeled as an amphipathic helix, there is no clear homology between the SIDs, as can be inferred from an alignment of the extended

Mad-SID with the SIDs of Pf1, Ume6, TIEG2 and BTEB1 (Figure 4.5(b)). Homology between these sequences is restricted to a short minimal core sequence, ϕ-ϕ-X-A-A-X-X-ϕ, with only one strictly conserved residue (Ala15). Thus, the extended sequence motif discriminates the Mad family proteins from other classes of PAH2-interacting proteins.

Furthermore, as residues that mediate important intermolecular interactions for the Mad proteins, such as Ile8 and the newly identified Glu23 and Met7, do not show any obvious similarities in Pf1, Ume6, TIEG2 or BTEB1, we suggest that these proteins show significant differences in their precise binding modes, either interacting via the minimal core only, or, analogous to the Mad proteins, involving additional residues outside this minimal motif to mediate intermolecular interactions. In fact, a recent study proposed a completely different model for the TIEG2-SID:PAH2 interaction based on molecular dynamics simulations and biochemical data (Pang et al., 2003). In this model, the TIEG2-SID is not wedged in the PAH2 four-helix bundle, but instead its helix is bound in a parallel manner to the outside surface formed by helix α1 and α2 by both hydrophobic and electrostatic interactions. Notably, the residues involved in crucial intermolecular contacts do not correspond to the minimal motif.

Conclusions

In conclusion, NMR data, binding studies, dynamics data, sequence alignment and previously described binding studies point to additional interactions between an extended Mad-SID and the PAH2 domain of mSin3. Furthermore, SID-Glu23 is involved in a key intermolecular interaction with PAH2^{B}-Lys165, previously attributed to SID-Glu20. This has led us to propose an extended sequence motif showing that residues outside the hydrophobic core of the SID interact with PAH2 and modulate binding affinity in a manner specific for the Mad-family.

Materials and Methods

Sample Preparation

All NMR studies were performed using a protein construct of 105 residues (denoted as PAH2^{B}) corresponding to residues 148-252 of the long variant of Mm. mSin3B (SpTrEMBL accession number Q62141) and a chemically synthesized polypeptide of 24 residues (denoted as SID24) corresponding residues 5-28 of Hs. Mad1 (SwissProt accession number Q05195). Cloning, expression and purification of the PAH2^{B} domain were done as described before (Spronk et al., 2001). Uniformly ^{15}N/^{13}C-labeled PAH2^{B} was prepared using [^{15}N]-NH$_4$Cl and [^{13}C$_6$]-glucose as sole nitrogen and carbon sources. NMR samples contained a 1.3 mM 1:1 complex of ^{13}C/^{15}N double-labeled PAH2^{B} and unlabeled SID24 in a buffer of 50 mM KH_2PO_4/K_2HPO_4 (pH 6.3) and $^{1}H_2O/^{2}H_2O$ (95% /5%), using the protease inhibitor Pefabloc and trace NaN_3 as preservative. The 1:1 stoichiometry of the PAH2:SID interaction was confirmed on the basis of both the results of titrating SID to an NMR sample of free PAH2 (Spronk et al., 2001) and the agreement of fit of the experimental SPR data with a 1:1 stoichiometry (data not shown). Samples with free PAH2^{B} and PAH2^{B} in complex with the 13-residue minimal SID were prepared as described before (Spronk et al., 2001). Constructs of Hs. Mad1 SIDs of different lengths were fused to the GB1-domain of streptococcal protein G as described earlier (Spronk et al., 2000) and were used in the SPR measurements.

Surface plasmon resonance interaction studies

SPR experiments were performed using a BIAcore 2000 biosensor instrument. Sensor chips and protein coupling chemicals were purchased from Biacore AB (Uppsala, Sweden). Proteins were coupled to the sensor chips by procedures recommended by the manufacturer using EDC and N-hydroxysucccinimide for activation of carboxylic groups of the sensor chips, followed by reaction of the protein ligand, and finally blocking the excess of activated groups by reaction with 1 M ethanol amine. The immobilization level was optimized for kinetic measurements, the maximal amount of bound analyte ($R_{\max}$) was kept low to minimize mass transport limitation effects on the binding curves and varied between 46 and 128 Resonance Units. Kinetic measurements were performed at 25 °C with a flow rate of 50 mL/min in HBS, pH 7.4.

Interaction of the PAH2 domain of mSin3B with the four Mad1/GB1 constructs was measured in duplicate in two independent manners. In the first approach, mSin3B was immobilized and the binding of the Mad1/GB1 constructs with variable Mad1-SID length was measured during 5 min at 10 different concentrations. Ligand and analyte were reversed in the second approach. Mad1/GB1-constructs were immobilized and the binding of 10 different concentrations of mSin3B was measured for 5 min. Regeneration of the sensor surface was performed with 50 mM HCl during 3 min. Interactions were also measured on sensor surfaces without immobilized protein as negative controls.

Equilibrium was reached within 30 s for all measured interactions. Steady-state binding kinetics was applied to calculate the binding constant of the interactions. At equilibrium, the relationship between response and concentration is defined as: $R_{eq} = K_A \cdot conc \cdot R_{max}/[1+K_A \cdot conc \cdot n]$. R_{eq} corresponds to the steady state value of a particular concentration of analyte, K_A is the association constant (in M^{-1}), R_{max} is the maximum amount of bound analyte and n is the number of independent ligand binding sites. The Kinetic Evaluation software of the manufacturer was used to generate overlay plots of 10 concentrations of analyte, to determine R_{eq} values for all concentrations corrected for the negative control responses, and to calculate the dissociation constant K_D (in M). Two independent values of the dissociation constants were calculated for the interaction of mSin3B with four Mad1-SID constructs. Average values are shown in Table 4.1.

NMR spectroscopy

All NMR experiments were collected at 20.0 °C on a sample with $^{13}C/^{15}N$ double-labeled PAH2^B and unlabeled SID24. The NMR experiments were carried out on Varian Unity Inova 500, 600, 750 and 800 MHz spectrometers. The data were processed using the NMRPipe suite (Delaglio et al., 1995) and analyzed using XEASY (Bartels et al., 1995). Assignment of PAH2^B was based on the assignment of PAH2^B in complex with the minimal SID of 13 residues (Spronk et al., 2000; Spronk et al., 2001). Confirmation of the assignments was obtained using 3D HNCA, HNCACB, CBCACONH, (H)CCH-TOCSY, ^{15}N NOESY-HSQC and aliphatic ^{13}C NOESY-HSQC spectra. Assignment of the extended SID was performed using the sequential assignment approach (Wüthrich, 1986) using a 2D $^{13}C/^{15}N$ double-filtered NOESY and a 3D doubly sensitivity enhanced $^{13}C/^{15}N$ filtered TOCSY-TOCSY (van Ingen, 2002), which was necessary to solve all overlap in the conventional 2D X-filtered TOCSY due to the high redundancy in amino acid composition.

Distance restraints for structure calculations were obtained from 3D ^{15}N NOESY-HSQC, aliphatic ^{13}C NOESY-HSQC, aromatic ^{13}C NOESY-HSQC, 2D $^{13}C/^{15}N$ double filtered NOESY experiments. Intermolecular NOEs were distinguished from intramolecular NOEs using a 3D (^{13}C-edited-$^{13}C/^{15}N$-filtered)-HMQC-NOESY and a (^{15}N-edited-$^{15}N/^{13}C$-filtered)-HSQC-NOESY. The NOE mixing time in all NOESY experiments was set to 100 ms. To determine changes in the ^{13}C, ^{15}N and ^{1}H chemical shifts of the side chain amide groups of Asn and Gln, HNCO experiments optimized for NH_2-moieties were recorded on samples containing the PAH2^B-SID13 and PAH2^B-SID24 complexes.

The ^{15}N-T_1, ^{15}N-$T_{1\rho}$ and $\{^1H\}$-^{15}N NOE experiments were recorded at 20.0 °C at 11.7 T. All experiments were recorded in an interleaved manner and in duplicate. Relaxation delays for the PAH2^B-SID24 complex were: 16, 96, 160, 256, 512, 640, 800 and 1024 ms for the T_1-experiment and 16, 32, 48, 80, 96, 112 and 128 ms for the $T_{1\rho}$-experiment. For the PAH2^B-SID13 complex, relaxation delays were: 16, 96, 256, 400, 608, 800 and 1024 ms for the T_1-experiment and 16, 32, 48, 64, 80, 96 and 128 ms for the $T_{1\rho}$-experiment. The $\{^1H\}$-^{15}N NOE values were derived from experiments using 3 s of either on- or off-resonance irradiation with a recycle delay of 1.5 s and an additional 10 s delay after each complex point.

Structure calculations and refinement

The NOE peak volumes were converted into distance ranges by normalization against peak volumes that were calibrated to known distances, and then overestimated 2-fold. Distance restraint ranges used lower limits of 0 Å and upper limits of 2.8, 3.5, 5.0 or 7.0 Å. Chemical shifts of C_α for PAH2 and H_α for the SID were used to derive a set of 140 dihedral angle restraints for the backbone angles ϕ and ψ using the program CSI version 2.0 (Wishart and Sykes, 1994). An initial set of 200 NMR structures was calculated using a Cartesian-space simulated annealing protocol in X-PLOR version 3.851 (Brünger, 1996). To improve local geometry and electrostatics, 106 lowest energy structures were then refined in water using a restrained molecular dynamics protocol under a CHARMM22 force field (Spronk et al., 2002). Briefly, the unrefined structures are solvated in a TIP3 water box, neutralized and energy minimized, followed by 3 ps of restrained MD. The initial temperature of the system is set to 300K and then cooled to 100 K. Force constants for experimental restraints are gradually lowered from 50 to 25 kcal mol^{-1} Å^{-2} and from 200 to 80 kcal mol^{-1} Å^{-2} for NOE and dihedral restraints, respectively. The resulting structures are energy minimized in 250 steps. Of these resulting structures, the 30 lowest energy structures with no distance restraint violations > 0.5 Å and dihedral angle violations > 5° were selected to form the final ensemble. Structures were analyzed using the programs PROCHECK-NMR (Laskowski et al., 1996) and WHAT-IF (Vriend, 1990).

Relaxation data analysis

Peak intensities were extracted from the spectra by fitting peaks using the nlinLS module of NMRPipe

(Delaglio et al., 1995). Errors in peak heights were estimated by the root-mean-square of the noise-level. The longitudinal and transverse relaxation rates were calculated by fitting the extracted intensities to a mono-exponential decay using the modelXY module of NMRPipe. Errors (1σ confidence limits) were estimated from 21 Monte Carlo simulations. In a conservative approach, all errors for R_1, $R_{1\rho}$ and NOE were estimated to be at least 4% to prevent falsely attributed contributions of internal motions or exchange. The global diffusion tensor was determined using the program TENSOR 1.1 (Dosset et al., 2000), resulting in an isotropic diffusion tensor with a global rotation correlation time of 11.2 ns for the PAH2^B-SID24 complex and 9.9 ns for the PAH2^B-SID13 complex. We used the lowest energy structure and a subset of the R_1 and $R_{1\rho}$ values, composed of residues that: *i*) do not overlap; *ii*) have NOE values >= 0.55 indicating that internal motion of these residues is limited; *iii*) have $R_{1\rho}$ within 1 SD from the mean value, indicating that chemical exchange contribution to $R_{1\rho}$ is absent or very small. Although, the $R_{1\rho}/R_1$-ratios of the extended SID complex suggest some anisotropy of the diffusion tensor, this effect could not statistically significantly be discriminated from an isotropic model. Subsequently, the R_1, $R_{1\rho}$ and NOE values served as input for the program Modelfree 4.15 (Mandel et al., 1995; Palmer et al., 1991) to extract parameters describing the internal motions in the model-free approach (Lipari and Szabo, 1982ab; Clore et al., 1990). After model selection, model parameters and the global rotation correlation time (τ_c) were optimized.

Data bank accession numbers

The coordinates of the structures were deposited in the RCSB Protein Data Bank under PDB accession number 1PD7. The chemical shifts of the extended SID and chemical shifts and ^{15}N relaxation data of the PAH2^B domain are deposited in the BioMagResBank under BMRB accession number 5457 and 5808, respectively.

Chapter 5.

Hugo van Ingen,
Maria A.H. Baltussen,
Jan Aelen
and Geerten W. Vuister

*Journal of
Molecular Biology*
(2006), **358**, p. 485-497

Role of structural and dynamical plasticity in Sin3: the free PAH2 domain is a folded module in mSin3B

Acknowledgements. We gratefully thank dr. John Lansing and prof. dr. P.E. Wright for the opportunity to use GLOVE for the analysis of the dynamics data and helpful discussions. We would like to thank dr. Mark Hilge for help with the structure calculations, dr. Sander Nabuurs for calculation of the *U*-factors and prof. dr. C.W. Hilbers for his comments on the manuscript. This research was funded by a "Jonge Chemici" grant of the Netherlands Organization for Scientific Research, NWO-CW (JC 99-03) to G.W.V.

The co-repressor Sin3 is the essential scaffold protein of the Sin3/HDAC co-repressor complex, which is recruited to the DNA by a diverse group of transcriptional repressors, targeting genes involved in the regulation of the cell cycle, proliferation and differentiation. Sin3 contains four repeats commonly denoted as Paired Amphipathic Helix (PAH1-4) domains that provide the principal interaction surface for various repressors. Here, we present the first structure of the free state of the PAH2 domain and discuss its implications for interaction with the repressors. The unbound conformation is highly similar to the conformation observed when bound to either the Mad1 or HBP1 repressor, suggesting that the PAH2 domain serves as a template that guides proper folding of the unstructured repressor. The free PAH2 domain shows µs-ms conformational exchange between the folded, major state and a partially unfolded, minor state. Upon complex formation, we observe a significant decrease in fast time scale flexibility of local regions of the protein, correlated with the formation of intermolecular contacts, and an overall decrease in the slow time scale conformational exchange. On the basis of our data and using a multiple sequence alignment of all PAH domains, we suggest that the PAH1, PAH2 and PAH3 domains form pre-folded binding modules in full length Sin3 like 'beads-on-a-string' and act as folding templates for the interaction domains of their targets.

Introduction

The regulation of gene expression in eukaryotes requires the formation of specific protein-DNA and protein-protein complexes (Roeder, 2003). Transcriptional activators and repressors recognize specific DNA sequences upstream of the promoter of target genes, resulting in either up- or down-regulation of transcription (Ptashne and Gann, 1997; Ptashne, 2005). This is achieved indirectly via recruitment of co-regulators and associated proteins that together are capable of shaping the chromatin structure into an active or repressive state (Struhl, 1997; Burke and Baniahmad, 2000; Berger, 2002; Cosma, 2002). The association between the activator/repressor and the co-regulator is a crucial step as it provides the link between target gene and chromatin modification.

A diverse group of repressors recruit a co-regulator complex containing the highly conserved co-repressor Sin3 (for a recent review, see Silverstein and Ekwall, 2005) and the Class I histone deacetylating proteins (HDAC1/2). The HDAC proteins deacetylate specific lysine residues in the tails of the histone proteins, resulting in a compact, repressive chromatin state (Hassig et al., 1997; Laherty et al., 1997; Zhang et al., 1997; Vermeulen et al., 2004; Eberharter et al., 2002). The core Sin3/HDAC complex further consists of Sin3 associated proteins (SAP18 and SAP30), that are thought to stabilize the complex (Zhang et al., 1997; Zhang et al., 1998), and histone targeting proteins (RbAp48 and RbAp46), that stabilize contact with the nucleosome (Hassig et al., 1997; Zhang et al., 1997). Additionally, the core

Sin3/HDAC co-repressor complex exists of SDS3 (Alland et al., 2002), SAP130 and SAP180 (Fleisher et al., 2003) and can be extended with extra catalytic modules, such as the monosaccharide *O*-GlcNAc transferase OGT (Yang et al., 2002).

Sin3 contains four imperfect repeats, denoted as PAH domains that were originally presumed to fold into paired amphipathic helices (Wang et al., 1990), and a highly conserved region that is responsible for interaction with HDAC1/2, referred to as the HDAC interaction domain (HID) (Figure 5.1(a)). The second PAH domain, PAH2, and to a lesser extent the PAH1 and PAH3 domains have been identified as the regions required for interaction with the repressor (for an overview, see Silverstein and Ekwall, 2005), whereas the PAH4 domain is required for association with the monosaccharide transferase OGT (Yang et al., 2002). The mammalian homolog of Sin3 exists in two isoforms, mSin3A and mSin3B (Ayer et al., 1995), that most likely have separable functional roles (Dannenberg et al., 2005; Cowley et al., 2005).

Structurally, Sin3 has remained largely uncharacterized. So far, the interaction between Sin3 and the transcriptional repressor Mad1 has received considerable attention, as Mad1 is part of the Myc/Max/Mad network involved in the switch between cell proliferation and cell differentiation (Grandori et al., 2000; Adhikary and Eilers, 2005). Solution structures of the PAH2 domain complexed to the Sin3 interacting domain (SID) of Mad1 have shown that the complex is folded as a ‘wedged helical bundle’, in which the PAH2 domain adapts a four helix bundle conformation into which the α-helix of the SID is inserted (Brubaker et al., 2000; Spronk et al., 2000; van Ingen et al., 2004). Recently, the HMG box-containing repressor HBP1, which targets several cell cycle- and differentiation-specific genes, was identified to interact with the PAH2 domain (Swanson et al., 2004). Interestingly, the structure of this complex showed a reversed orientation of the SID relative to the Mad1-complex while maintaining the overall fold. The reversal of this helix orientation is correlated with a reversal in the SID sequence motif (Swanson et al., 2004).

In contrast to the well-studied bound state of the PAH2 domain, the unbound state of the domain has not been described. However, knowledge of both bound and unbound states is paramount to allow an evaluation of the structural and dynamical changes induced upon complex formation and will result in a better understanding of the molecular mechanism of repressor-co-repressor interaction. Here, we report the solution structure and dynamics on fast and slow time scales of the unbound PAH2 domain of mammalian Sin3B and discuss its implications for the PAH-SID interaction. Previous studies have suggested that the PAH2 domain might not be fully folded in its unbound form and that PAH2 and the SID experience a coupled folding transition. We show that the domain is folded and that the structural and dynamical plasticity in PAH2 upon complex formation is limited to a fine-tuning of the interaction surface. In addition, we show that the PAH1-3 domains most likely form folded modules within Sin3, as ‘beads-on-string’, and are ready to accommodate their specific interaction partners.

Results

Optimization of the PAH2 domain construct

The previously determined structures of the PAH2 domain of mSin3B in complex with a Sin3 Interacting Domain (SID) of Mad1 used a PAH2 domain construct of 105 residues and showed that the folded part comprises about 75 residues (Spronk et al., 2000; van Ingen et al., 2004). For the study of the free PAH2 domain a construct was made consisting of 85 amino acids (residues 148-232 of Mm. mSin3B, denoted as PAH2^{B}-85), comprising the domain with 5 extra

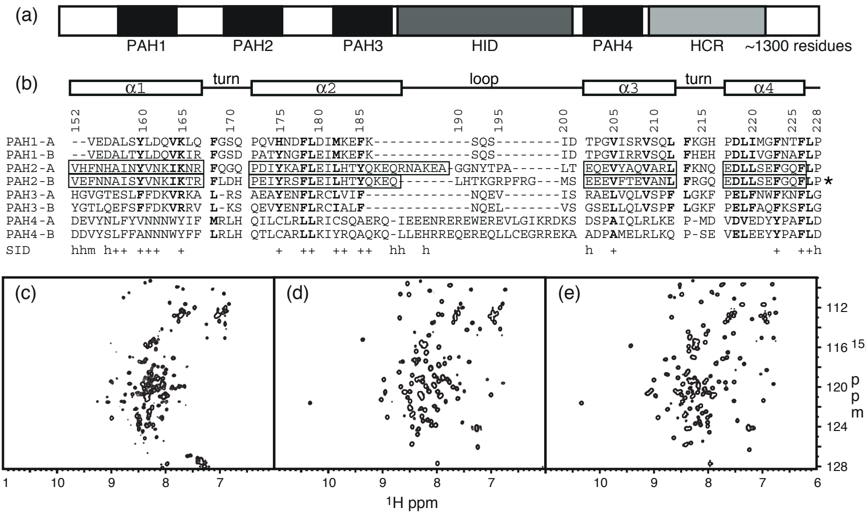


Figure 5.1. (a) domain organization of Sin3 (PAH = poly amphipathic helix domain; HID = histone deacetylase interacting domain; HCR = highly conserved region). (b) structure based sequence alignment of the PAH domain of mouse mSin3A and mSin3B. Secondary structure is indicated in the top and for the two experimentally determined structures by the black boxes. Residues in bold type are conserved or conservatively mutated in the PAH1-3 domains. Sequence used in this study is indicated with an asterisk. Residues that engage in intermolecular interaction with the Sin3 Interacting Domain (SID) of Mad1 (m), HBP1 (h) or both (+) are indicated in the bottom row. (c, d, e) ^{15}N-HSQC NMR spectra of the PAH2^{B} domain using a 85 residue construct (c; res. 148-232), a 105 residue construct, extended at the C-terminus (d; res.148-252) and the 105 residue construct in complex with the SID of Mad1 (e).

residues at both the N- and C-terminus. To assess the conformational integrity of this domain, a 2D ^{1}H-^{15}N correlation spectrum (^{15}N-HSQC) was recorded (shown in Figure 5.1(c)). The lack of dispersion of backbone amide resonances and the severe line broadening of several resonances indicate that this polypeptide does not adopt a well-defined tertiary structure in solution. The observed conformational heterogeneity persists over a wide temperature range and under different buffer conditions.

In sharp contrast, the HSQC spectrum obtained from a PAH2 construct extended with an additional 20 residues at the C-terminus (residues 148-252 of Mm. mSin3B, denoted as PAH2^{B}-105) shows a greatly improved spectrum (Figure 5.1(d)). This improvement is probably caused by an enhanced solubility of the longer construct in combination with reduced aggregation (see also the Discussion). The increased dispersion and the more homogeneous linewidths suggest that the PAH2^{B} domain is folded. In order to assess potential interference between the PAH2 domain and its C-terminal tail, we analyzed both the structural and dynamical properties. Both $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ chemical shifts of all residues are within 1 standard deviation from random coil values. Only intra-residual and sequential NOE cross peaks could be identified in the NOESY spectra (data not shown). Furthermore, fast time scale dynamics indicates that the tail is extremely flexible with local motions on a time scale of about one-tenth of the global tumbling rate. Relaxation-dispersion measurements also did not give any indication for the existence of a minor, possibly interfering, conformation for this region (*vide infra*).

In addition, the tail does not impede binding of the SID. The spectral dispersion in the

HSQC spectrum of PAH2^B-105 complexed to the Mad1-SID is further improved (Figure 5.1(e)). The overall resonance pattern does not change drastically between unbound and bound forms of PAH2^B-105, which suggests that the fold of PAH2 is essentially preserved between unbound and peptide-bound states. Hence, we conclude that the C-terminal tail is unstructured in our construct and does not interact with the binding pocket of PAH2. Given its superior stability and spectral quality, this PAH2^B-105 construct was used for further structural and dynamic studies.

Structure

The backbone traces of the unbound PAH2^B-105 ensemble and the best representative structure are shown in figures 5.2(a) and 5.2(b), respectively. In total 1007 unique distance restraints and 131 ϕ,ψ dihedral angle restraints were used to calculate the solution structure. Residues 233-252 that constitute the unstructured C-terminal tail, were not included in the structure calculation. The pairwise heavy atom RMSD of the ordered parts of the structure is 0.68 Å for the backbone and 1.46 Å overall, indicating a well-defined ensemble (Table 5.1). Analysis of the per-residue quality parameter *U* shows that the structural quality of the ensemble of structures is well-determined by the experimental data (see Figure 5.7A in the Appendix; Nabuurs et al., 2005). Further statistical data for the ensemble are reported in Table 5.1.

The domain adopts a 4-helix bundle fold that was found also for the structure of the complex of PAH2 and Mad1-SID (Brubaker et al., 2000; Spronk et al., 2000; van Ingen et al., 2004). The four helices are formed by

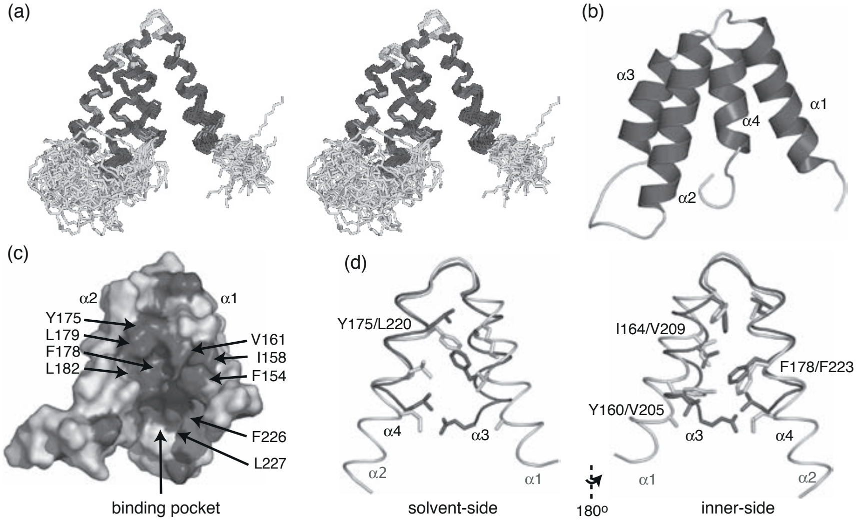


Figure 5.2. (a) wall-eye stereo view of the backbone traces of residues 148-232 of the ensemble of 30 structures of the free PAH2^B-105 domain. Residues 233-252 in the unstructured C-terminal tail were excluded from the structure calculation (b) cartoon representation of the best representative structure of the ensemble. Helices are shown in dark gray and are labeled α1 to α4; the turns and coil regions are shown in light gray. (c) surface representation of unbound PAH2^B, showing the binding pocket Hydrophobic residues are shown in dark gray. Hydrophobic residues in the binding pocket are labeled. (d) comparison of the equivalent inter helical angles α1,α2 (light gray) and α3,α4 (dark gray). Helix α2 and α4 and the adjacent turns are superimposed to show differences in packing in the helix interface (residues shown in stick representation). Residues crucial for the inter helical angle are labeled (α1-2 /α3-4).

Table 5.1.
Structural statistics for the PAH2^{B} domain[a]

A. Restraint information	
Total number of distance restraints	1007
intra-residual/sequential/medium/long	408/251/228/120
TALOS derived dihedral angle restraints ϕ/ψ	65/66
B. Average RMS deviation from experimental restraints	
All distance restraints (Å)	0.035 ± 0.002
All dihedral angle restraints (o)	0.26 ± 0.07
C. Pairwise Cartesian RMS deviation (Å)	
Ordered backbone heavy atoms	0.68 ± 0.12
Ordered all heavy atoms	1.45 ± 0.18
Global backbone heavy atoms	2.66 ± 0.62
Global all heavy atoms	3.48 ± 0.52
D. Ramachandran quality parameters (%)	
Residues in most favoured regions	86.2
Residues in allowed regions	12.8
Residues in additionally allowed regions	0.5
Residues in disallowed regions	0.5
E. Abnormalities found in structural checks	
Abnormally short interatomic distances	3 ± 1
Unsatisfied H-bond acceptors (buried)	0 ± 0
Unsatisfied H-bond donors (buried)	3 ± 2
F. Average RMS deviation from current reliable structures (RMS Z-scores, null deviation = 1)	
Bond lengths	0.72 ± 0.02
Bond angles	0.73 ± 0.02
Omega angle restraints	0.62 ± 0.05
Side-chain planarity	0.72 ± 0.09
Improper dihedral distribution	0.74 ± 0.03
Inside/outside distribution	1.07 ± 0.04
G. Average deviation from current reliable structures (Z-scores, null deviation = 0)	
2nd generation packing quality	-0.35 ± 0.21
Ramachandran plot appearance	-2.1 ± 0.7
Chi-1/Chi-2 rotamer normality	-0.62 ± 0.55

[a] statistics are given for residues 148-232 of PAH2^{B}-105, excluding the C-terminal tail. Ordered regions are residues 158-189 and 202-226.

residues 153-167 (α1), 172-189 (α2), 202-212 (α3) and 217-226 (α4). Residues 168-171 and 213-216 fold into turns and residues 190-201 form a large loop between helices α2 and α3. The most striking feature of the PAH2 structure is its large solvent exposed hydrophobic binding pocket of ca. 510 Å^{2} (averaged over the ensemble) (cf. Figure 5.2(c)). This exposed conformation is caused by poor packing of helix α1 to the remainder of the helical bundle. Whereas the helices α2-α3 and α3-α4 arrange as nearly anti-parallel, the angle between α1 and α2 is ~145°. As a result, tight packing of helix α1 and α2 is restricted to two helical turns adjacent to the connecting turn. In contrast, helices α3 and α4 pack closely along their full lengths (Figure 5.2(d)). The interface between the latter two helices is composed of complementary bulky (F206, L220, F223) and smaller side chains (V205, V209, A210, G224).

In contrast, a large angle between α1 and α2 is required to accommodate the bulky side chains in their interface (Y160, I164, K165, Y175, F178, L179). Most notably, the bulky F223 in α4 faces the smaller V205 and V209 in α3, while the corresponding residue F178 in α2 points towards the also bulky residues Y160 and I164 in α1. Furthermore, Y175 in the top of the α1,α2-interface is more bulky than the corresponding L220 in the α3,α4-interface (Figure 5.2(d)). The resulting poor packing of helices α1 and α2 creates a gap, which is further emphasized as these helices are longer than their counterparts α3 and α4.

To examine possible structural changes of the PAH2^B domain upon binding of target peptides, we compared the structure of unbound domain with the two different complexes solved to date, the PAH2-Mad1 (Brubaker et al., 2000; Spronk et al., 2000; van Ingen et al., 2004) and PAH2-HBP1 complex (Swanson et al., 2004). A structural superposition of the best representatives of each ensemble is shown in Figure 5.3. The well-defined parts of the helices in the ensemble of structures of the free domain and the Mad1-complex superimpose well with an RMSD of 1.09 Å (residues 158-186;202-223), indicating that the global fold of PAH2 is unchanged after ligand binding. For the HBP1-complex, a slightly higher RMSD of 1.37 Å is found for the same region. Excluding helix α1, the RMSD drops to 0.86 Å, reflecting a small but noticeable difference in the angle between helices α1 and α2 in the two ensembles of structures, which was also reported in the study of Swanson et al, 2004.

To compare the side chain orientations between the three structures, only residues with sufficient experimental restraints to result in well-defined side chain orientations in the three ensembles of structures were taken into account. The resulting heavy atom RMSD is 1.40 Å (1.61 Å) between the free structure and the Mad1-complex (HBP1-complex). This is roughly within the precision of the ensembles. More specifically, crucial residues in the binding pocket have nearly identical side chain positions in both the free and bound forms (cf. Figure 5.3(b)).

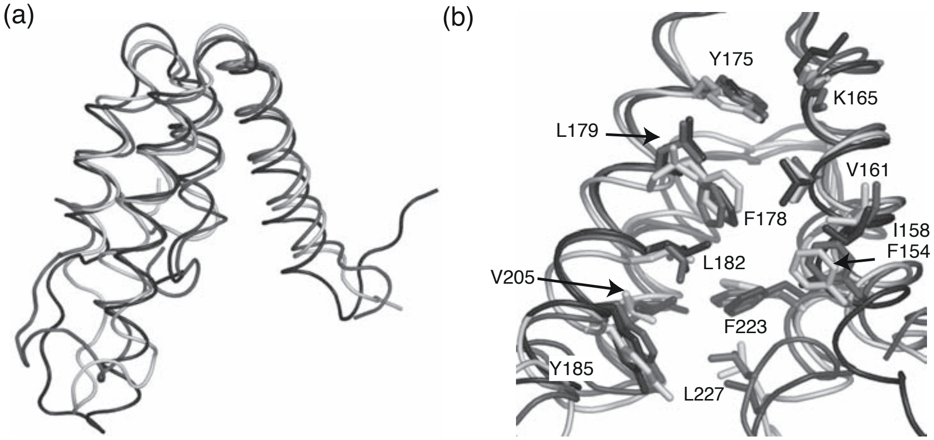


Figure 5.3. (a) superposition of the backbones of the unbound PAH2^B structure (dark gray), the PAH2^B domain bound to Mad1-SID24 (light gray; PDB code 1PD7) and PAH2^A bound to HBP1-SID (black; PDB code 1S5R). (b) enlarged view of the superposition displayed in (a), showing the side chains of residues that interact with the Mad1-SID and HBP1-SID. Color code as in (a), residue labeling corresponds to PAH2^B.

In general, these residues form the 'floor' of the binding pocket. Several residues at the edges of the pocket, e.g. residues E183, Y160, H183, Q186, Q225 and F226, are disordered in the free structure. Interestingly, residue F154 has a well-defined side chain orientation in both the unbound state and the Mad1-SID complex, but is disordered in the HBP1 complex. In the Mad1-complex it is part of the binding pocket and is responsible for several intermolecular contacts, while in the HBP1-complex it is pointing away from the binding pocket, into the solvent.

Fast time scale dynamics

In addition to inducing structural changes, binding events usually affect the dynamical behavior of proteins. In earlier work, we have characterized in detail the backbone dynamics of the Mad1-SID complex occurring on ps-ns time scale (van Ingen et al., 2004). Here, we also investigated these fast backbone motions in the unbound state of PAH2^B. The ^{15}N longitudinal and transverse relaxation rates (R_1 and $R_{1\rho}$, respectively) and the steady state Nuclear Overhauser Effect ($\{^1H\}$-^{15}N NOE) were measured for 69 out of 102 non-proline residues (cf. Figure 5.8A, in the Appendix). The resonances of the remaining 33 residues were overlapping too severely to extract reliable peak intensities. The structured regions have NOE values close to 0.8, indicating that fast local motion is limited. Low NOE values with concomitant low $R_{1\rho}/R_1$ ratios indicate high local flexibility for the N-terminus, the loop between helices α2 and α3 and for the C-terminal tail. A few residues have high $R_{1\rho}/R_1$ ratios, indicating the presence of chemical exchange on µs-ms time scale. This is elaborated on in more detail in the slow dynamics section.

The C-terminal tail is highly flexible on a fast time scale, judged from its very low $\{^1H\}$-^{15}N NOE values (average 0.02 ± 0.47) and $R_{1\rho}/R_1$ ratio's (average 3.4 ± 2.3 s^{-1}). It has been shown that the presence of such a flexible tail frustrates the quantitative description of dynamics using the model-free approach, as evidenced by anomalous high values for the model-free parameters (Viles et al., 2001). As the reorientation of the tail is on the time scale of the global reorientation, this results in a time dependence of the diffusion tensor which constitutes a violation of the central assumption of the model-free approach that local and global motions can be decoupled (Lipari and Szabo, 1982ab). In contrast, reduced spectral density mapping does not make any assumption about the nature or time-scales of local and global motions and is used to qualitatively describe the local dynamics in term of the spectral densities $J(\omega)$ at the frequencies 0, ω_N and $0.87\omega_H$ (Farrow et al., 1995; Lefèvre et al., 1996; Ishima and Nagayama et al., 1996). In short, assuming no internal motions and a rigid isotropic rotor, $J(0)$ equals τ_c, whereas smaller values for $J(0)$ indicate internal mobility of the N-H bond on a sub-nanosecond time scale. Fast internal motions also result in an increase in $J(0.87\omega_H)$ values while the $J(\omega_N)$ spectral density decreases for small to medium sized proteins. Motions on a micro-millisecond time scale result in a chemical exchange contribution to the transverse relaxation rate. When the observed $R_{1\rho}$ is not corrected for this additional contribution, these motions are reflected by elevated values for $J(0)$, but neither affect the $J(\omega_N)$ nor $J(0.87\omega_H)$ values.

Figure 5.4(a) presents $J(0.87\omega_H)$-$J(0)$ and $J(\omega_N)$-$J(0)$ correlation plots and illustrates the markedly different dynamics of the core and the tail of the protein. Three clusters can readily be identified. These clusters are schematically represented on a structure model that includes the C-terminal tail (Figure 5.4(b)). Firstly, the disordered C-terminal and N-terminal tails (group I; residues 148-151, 238-252) are extremely flexible, as indicated by the very low $J(0)$ values with corresponding high $J(0.87\omega_H)$ values and lower $J(\omega_N)$ values.

The $J(0.87\omega_H)$ and $J(0)$ values and the $J(\omega_N)$ and $J(0)$ values show the expected linear

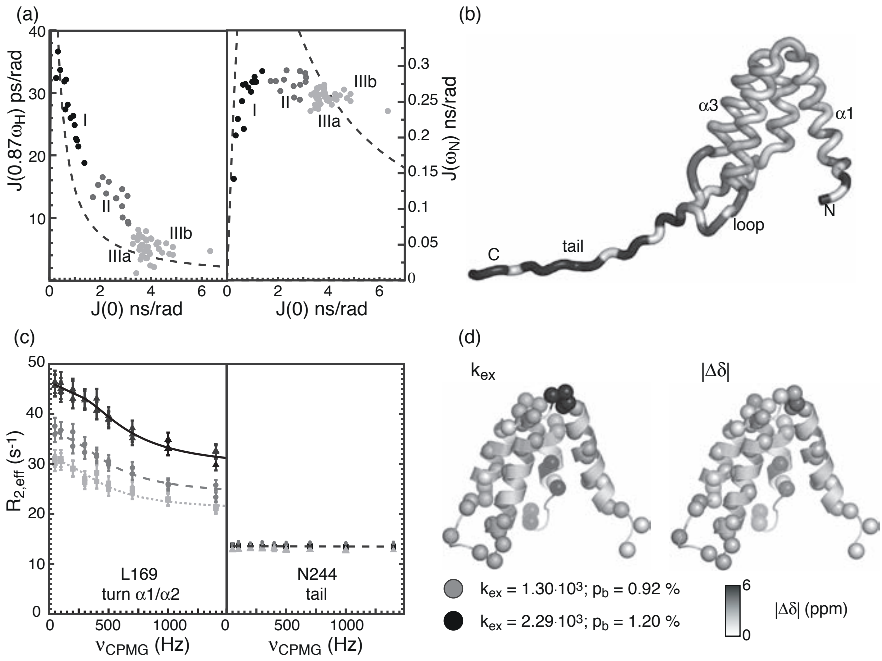


Figure 5.4. Results from the fast (panels (a) and (b)) and slow (panels (c) and (d)) dynamics analysis. (a) Results of the reduced spectral density mapping analysis. Spectral densities $J(0.87\omega_H)$ (left panel) and $J(\omega_N)$ (right panel) are plotted as function of the corresponding apparent $J(0)$ value (uncorrected for chemical exchange). The solid curves show the expected correlation for a rigid isotropic rotor. Data points cluster in three groups, labeled I (very flexible; black), II (flexible; dark gray) and III (rigid; light gray). Group III can be divided in a subgroup without and with significant chemical exchange contribution, indicated with IIIa and IIIb, respectively. (b) The three groups are indicated on a structure-model including the C-terminal tail. Residues with no available data are color coded very light gray. (c) representative ^{15}N R_2 relaxation-dispersion curves at 500 (light gray; duplicate), 600 (dark gray; duplicate) and 800 MHz (black; triplicate) proton frequency. The solid lines in the left panel are the best fit curves obtained from a global fit as described in Materials and Methods; a broken straight line is shown in the right panel to guide the eye. (d) Results from the dispersion data analysis. Values of the exchange rate constant (k_{ex}; left) and the magnitude of the chemical shift difference ($|\Delta\delta_{ba}|$, right) are plotted on the structure. The backbone nitrogen nuclei are shown as spheres for residues for which dispersion data was acquired. Residues without significant dispersion are color coded very light grey in the left panel.

correlation, indicating the superposition of a global and local motion (Lefèvre et al., 1996).

Secondly, the relatively disordered first two turns of helix α1, the α2-α3 loop and the transition region between the core of the protein and the tail (group II, residues 152-156, 190-200, 234-237) also have increased flexibility on a sub-nanosecond and picosecond time scale as judged by their lower $J(0)$ and higher $J(0.87\omega_H)$ spectral densities. Group III (residues 157-189, 201-233) contains all residues in structured regions (except the N-terminal part of helix α1) and can be divided in two subgroups. Group IIIa has a rather homogeneous $J(0)$ close to the theoretical curve for a rigid isotropic rotor, reflecting restricted internal motions for the N-H vector. The apparent isotropic τ_c is 9 ns for this group. The spread in $J(\omega_N)$ and $J(0.87\omega_H)$ values reflects a small anisotropy in the rotational diffusion. Group IIIb has elevated apparent $J(0)$ values, indicating significant chemical exchange contributions.

These residues (164, 167-169, 174, 183, 201, 215, 223, 226 and 232) were also identified by the relaxation-dispersion experiments and are further discussed below

Slow time scale dynamics

To probe the backbone dynamics at a μs-ms time scale, effective ^{15}N-R_2 values were determined as a function of pulse spacing in CPMG relaxation dispersion experiments. Dispersion profiles could be determined for 68 residue at three magnetic field strengths. Residues in the N-terminus and the C-terminal tail did not show any dispersion of their effective R_2 values. Most residues in the core of the protein and in the transition region between helix α4 and the tail showed significant R_2 dispersion. Representative dispersion profiles are shown in Figure 5.4(c).

Assuming two-site exchange between a major (a) and minor (b) state, exchange parameters were determined by global fitting of the Carver-Richards-Jen all time scale approximation to the dispersion profiles. All residues are in fast-to-intermediate exchange. Out of 43 residues with significant dispersion, 33 residues could be fitted in a single cluster with exchange rate, k_{ex}, equal to $1.30 \cdot 10^3 \pm 0.05 \cdot 10^3$ s^{-1} and the population of the minor state, p_b, equal to 0.92 ± 0.05% . The remaining 10 residues were fitted using a second cluster with k_{ex} equal to $2.29 \cdot 10^3 \pm 0.13 \cdot 10^3$ s^{-1} and p_b equal to 1.20 ± 0.07% (see Table 5.3A in the Appendix).

The two clusters are mapped onto the unbound PAH2^B structure (left panel of Figure 5.4(d)). The first cluster contains mostly residues in helices and a few in the α2-α3 loop. The cluster with the higher exchange rate comprises residues in the turn between helix α1 and α2 and the C-terminal part of α4 with subsequent residues. Relaxation dispersion profiles also depend on the magnitude of the frequency difference between the two exchanging states a and b ($|\Delta\omega_{ba}|$). This frequency difference is more conveniently expressed as a difference in chemical shifts, $|\Delta\delta_{ba}|$. The right panel of Figure 5.4(d) displays the $|\Delta\delta_{ba}|$ values mapped onto the PAH2^B structure. On average, the chemical shift difference is circa 2 ppm. Residues in helix α3 experience on average the smallest change in chemical shift (1.38 ± 0.3 ppm), while three regions form 'hot-spots' with large average values for $|\Delta\delta_{ba}|$: the turn between α1 and α2 (3.4 ± 2 ppm), residues 198-201 that form a relatively ordered part of the loop adjacent to helix α3 (3.1 ± 1 ppm) and the C-terminal part of helix α4 with adjacent residues in the linker region (4.6 ± 2 ppm).

Structural characterization of the minor conformation is possible only if both the magnitude and the sign of $\Delta\delta_{ba}$ is known, allowing reconstruction of the chemical shifts of the minor state (Skrynnikov et al, 2002; Grey et al., 2003; McElheny et al., 2005; Beach et al., 2005). We determined the sign using a new experiment, named CEESY (see Chapter 6) and confirmed these results using the comparison of peak positions in heteronuclear single quantum and multiple quantum correlation spectra (Skrynnikov et al., 2002). The resulting signs of the chemical shift

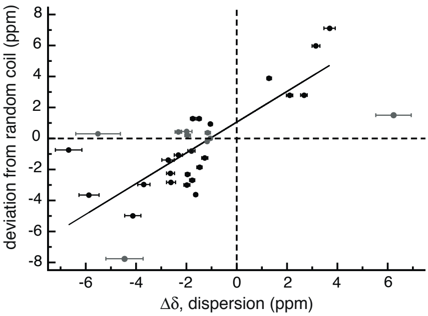


Figure 5.5. Correlation between the signed chemical shift differences between major and minor forms of the PAH2^B domain and the chemical shift difference with the random coil shift. The solid line is the linear squares best fit line through the data points ($R = 0.82$). Residues for which the relative error on the effective R_2 was larger than 10% and residues for which the deviation from the random coil shift was less than 0.5 ppm (shown in gray) were excluded from the fit.

Table 5.2
Structural interpretation of exchange process.

residue	HN/N[a]	sign $\Delta\delta_{ba}$[b]	$\Delta\delta_{ro}$[c]	random coil[d]
F168	HN	n.d.		-
	N	+	+7.11	yes
L169	HN	+	+1.03	yes
	N	−	−3.65	yes
D170	HN	−	−0.12	yes
	N	+	+5.97	yes
H171	HN	n.d.		-
	N	−	+0.92	no
R198	HN	n.d.		-
	N	−	−2.97	yes
G199	HN	n.d.		-
	N	−	−2.34	yes
M200	HN	n.d.		-
	N	−	−0.80	yes
S201	HN	−	−0.52	yes
	N	−	−4.98	yes
F223	HN	+	+0.46	yes
	N	−	−0.74	yes
K231	HN	+	+0.39	yes
	N	n.d.		-
R232	HN	+	+0.81	yes
	N	+	+1.50	yes

[a] sign information on exchange process for amide proton (HN) or nitrogen (N). Only residues in the 'hot-spots' are listed.

[b] sign of chemical shift difference between minor (b) and major (a) conformational state (n.d. : under threshold, sign not determined).

[c] difference with sign between the random coil (r) and the observed chemical shift (o).

[d] 'yes' if the minor state is more random coil-like ($\Delta\delta_{ba}$ and $\Delta\delta_{ro}$ have the same sign), 'no' if minor state is less random coil-like ($\Delta\delta_{ba}$ and $\Delta\delta_{ro}$ have opposite signs).

differences were combined with their magnitude, as determined from the relaxation-dispersion experiments. Comparison of the resulting signed chemical shift differences, $\Delta\delta_{ba}$, with the chemical shift differences between the free and bound forms of PAH2 showed no correlation (R = 0.08). However, comparison of $\Delta\delta_{ba}$ with the deviation from the random coil shifts (Schwarzinger et al., 2000) resulted in a reasonable correlation (R = 0.82) (Figure 5.5), suggesting that the minor conformation is a (partially) unfolded state. Comparison with shifts of fully unfolded PAH2 in 8M guanidinium chloride solution resulted in a similar correlation (data not shown). Furthermore, our CEESY experiment enabled us to determine the sign of the frequency difference for the amide proton (see Chapter 6). As shown in Table 5.2, nearly all signs of $\Delta\delta_{ba}$ for both the amide proton and nitrogen in the three hot-spots indicate a shift toward random coil values, in agreement with an order-disorder transition.

Discussion

Influence of the C-terminal tail

We observed significant differences in the ^{15}N-HSQC spectra of two different PAH2^{B} domain constructs. While a spectrum consistent with one dominant folded conformation is observed for PAH2^{B}-105, that of PAH2^{B}-85 is typical for a protein in intermediate exchange between multiple conformations. Given the limited chemical shift dispersion, it is likely that the exchange process in PAH2^{B}-85 is an interconversion between the folded state and a partially unfolded molten globule state, similar to that in PAH2^{B}-105. There is no evidence that the observed stabilization of the folded conformation in PAH2^{B}-105 is caused by a (partial) association of the tail with the binding pocket, or any other part of the PAH2 domain. Given the extreme flexibility of the tail, transient interactions are also unlikely. Improvement in the solubility by the highly hydrophilic tail could contribute to the observed stabilization, although the short construct did not show severe problems with precipitation. Alternatively, it is plausible that the tail reduces the formation of soluble micro-aggregates that could be formed by partially unfolded states of the protein domain. In addition, charged residues in the tail could be involved in stabilizing long-range electrostatic interactions. Together, this could result in a stabilization of the folded conformation, resembling the way in which the GB1 domain can be used as a solubilization enhancement tag (Zhou et al., 2001).

Sequence analysis shows that the sequence of this C-terminal region is conserved in the B-isoforms of the PAH2 domain of the vertebrates human, mouse and the African clawed frog, which could point to a general requirement for this region for the stabilization of the PAH2 domain of mSin3B (data not shown). After submission of our manuscript, the solution structure of the PAH1 domain of mSin3B bound to the SID of the NRSF/REST repressor was published (Nomura et al., 2005). Interestingly, this structure shows that a conserved region after the C-terminus of the PAH1 domain is folded as a short helix and is packed against helices α3 and α4 of the PAH1 domain (Nomura et al., 2005).

Structural plasticity of complex formation

The global folds of the unbound, Mad1-bound and HBP1-bound PAH2 domain are virtually identical. This is rather surprising, as the hydrophobic interaction surface of the PAH2 domain is exposed to the solvent in the unbound form (cf. Figure 5.2(c)). This 'open' conformation is a consequence of the poor complementarity of side chains in the α1,α2-interface, forcing helix α1 out of the core formed by helices α2, α3 and α4 (Figure 5.2(d)). The limited structural rearrangement of the main chain, is paralleled by only minor changes in side chain conformation upon binding. Specifically, most side chains that directly interact with the SID have practically identical conformations in the unbound and both the Mad1-SID and HBP1-SID bound states.

In contrast, larger structural changes are thought to occur within the SID. Both the SID of Mad1 and HBP1 are most likely unstructured in their free state, but fold to form an amphipathic α-helix when bound to PAH2 (Eilers et al., 1999; Swanson et al., 2004). The folded SID interacts with the binding pocket of PAH2 via a precise arrangement of bulky and small side chains, like 'knobs-in-holes'. Interestingly, HBP1 and Mad1 bind with opposite helical orientations to complement the pre-defined 'knobs and holes' of PAH2, as the crucial SID residues are reversed in the sequence (Swanson et al., 2004). A recent mutation study stressed the importance of the complementarity of bulky and small residues in the center of the interaction surface as a basis for specificity and affinity of the SID-PAH2 interaction (Cowley et al., 2004). In addition to these general interactions, both Mad1 and HBP1 each have a number of unique interactions with PAH2, involving residues at the edge of the binding pocket. The unique stacking interaction between F154 of PAH2 and Y18 of Mad1 was identified as the critical determinant of specificity of Mad1 for the PAH2 *vs.* the PAH1 domain (le Guezennec et al., 2004). Notably, F154 becomes disordered in the HBP1-complex, as the HBP1-SID does not provide an aromatic residue at the corresponding position of Mad1-Y18. Conversely, V152, E153, P228 and a few residues at the C-terminal end of helix α2 uniquely interact with HBP1. In addition, residues at the edges of the binding pocket of PAH2 are involved in specific intermolecular interactions with residues outside the minimal Mad1-SID (van Ingen et al., 2004).

The view emerging from the current data indicates that complex formation involves a folding transition of the SID while the PAH2 domain experiences only limited structural rearrangements. Recent computational analyses of protein-protein interactions have shown that the conformation of crucial residues in unbound and bound state can be very similar for at least one of the interacting proteins (Li et al., 2004; Ramajani et al., 2004). The structured, unbound PAH2 domain might then serve as a folding template, defining the proper fold of the SID. This suggests that the short construct, PAH2^{B}-85, which is only partially folded, will have a lower binding affinity for the SID. A reasonable mechanism for complex formation is a combination of induced fit and conformational selection (Ramajani et al., 2004). In this model, a few residues of the SID serve as ‘anchors’ whose side chains are inserted into the well-defined pocket in the core of the interaction surface of PAH2, producing a native-like encounter complex. We suggest that the two central bulky hydrophobic residues (ϕ) in the ϕXXϕϕXAA minimal interaction motif, i.e. L12 and L13 in Mad1, act as anchor-residues. Binding affinity is subsequently optimized by a induced fit process in the periphery of the binding interface (Ramajani et al. 2004; Smith et al., 2005), including the formation of SID-dependent, unique interactions.

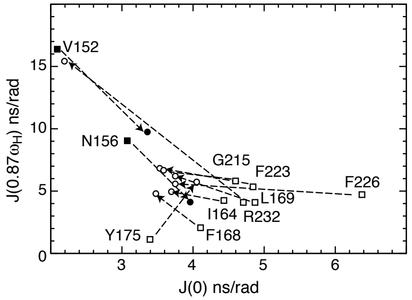


Figure 5.6. Comparison of the dynamics of unbound and Mad1-SID bound PAH2. The 10 most significant changes in the $J(0)$, $J(0.87\omega_H)$ correlation are shown. Squares represent data points of the unbound PAH2 domain, circles are shown for the Mad1-SID complex. Arrows are shown to connect residues in the free and bound state. Data points are labeled with their residue number. Filled symbols are used for residues in the first turns of helix α1, open symbols for residues in the rigid part of the structure.

Dynamical plasticity of complex formation

The PAH2^B domain shows a rich and complex dynamics, both on the fast and slow time scale. Reduced spectral density mapping showed that the helices, the loop and the tail form three regions with increasing fast local flexibility. On a μs-ms time scale, the PAH2^B domain is interconverting with a minor form that is populated for roughly 1%. The turn between helices α1 and α2, a relatively ordered part of the loop and the C-terminal part of helix α4 with the adjacent linker region form three ‘hot-spots’ that experience large changes in chemical environment. The minor state most likely constitutes a partially unfolded state, as indicated by a shift towards random coil values for the chemical shifts of both the amide ^{1}H and ^{15}N nucleus. Interestingly, residues in the turn between helices α1 and α2 show exchange on a significantly faster time scale than the helices, which could be the result of an modulation of the angle between α1 and α2, altering the accessibility of the binding pocket .

Figure 5.6 shows the 10 most significant changes in the fast time scale dynamics between unbound PAH2^B and the Mad1-SID complex (van Ingen et al., 2004). Complex formation results in a reduction in fast time scale dynamics for the first two turns of α1, as shown by the drastic reduction in the $J(0.87\omega_H)$ density with a concomitant increase in $J(0)$. This correlates with the formation of intermolecular interactions which ‘lock’ this helix. Other parts of the interaction surface do not experience such large changes, indicating that the impact of binding on the fast time scale dynamics seems to be restricted to that part of the interaction surface that is more flexible in the unbound state. On the other hand, binding reduces slow time scale motions for all residues that experience chemical exchange in the unbound state, as demonstrated by a decrease in their apparent $J(0)$ values. This finding was also confirmed by the absence of significant dispersion of ^{15}N transverse relaxation rates, as determined with relaxation dispersion experiments on the PAH2-Mad1 complex (data not shown). Interaction with the SID thus stabilizes the PAH2 domain with respect to the partially unfolded state, which may be correlated with the protection of the hydrophobic binding pocket from the solvent. Finally, the change in dynamics of residue Y175 seems to signify a more isotropic global rotational diffusion for the complex compared to the free domain (Figure 5.6).

PAH2 in Sin3A *vs.* Sin3B

The vertebrate Sin3 protein exists in two isoforms, known as mSin3A and mSin3B (Ayer et al., 1995), as a result of gene duplication early in evolution (Silverstein and Ekwall, 2005). Interestingly, NMR spectra of the unbound PAH2 domain of mouse Sin3A, $PAH2^A$, showed that the domain exists in two forms in a roughly one-to-one ratio that interconvert slowly on the NMR time scale (Brubaker et al., 2000; Swanson et al., 2004), as opposed to the fast-to-intermediate interconversion with a 1% populated minor state observed for $PAH2^B$ in this study. We suggest that both exchange processes are interconversions between a partially unfolded and a fully folded state. The observed differences in the kinetics and thermodynamics of this interconversion have to originate from differences in sequence, as the experimental conditions for the $PAH2^A$ and $PAH2^B$ studies are comparable. Indeed, the 'hot spots' of exchange are also the regions with the highest sequence variability (cf. Figure 5.1(b)). Different thermodynamic stability of the PAH domain could be related to functional differences between mSin3A and mSin3B by altering the free energy of interaction with binding partners. Also, different life times for the folded form, caused by the different exchange time scales, can affect binding of interacting proteins.

The PAH domains in full length Sin3

To assess the conservation of crucial residues between the different PAH domains, we performed a multiple sequence alignment of 35 Sin3 homologues and subsequently aligned their PAH domains using ClustalW (Thompson et al., 1994; see Figure 5.9A in the Appendix). The region originally identified as the PAH4 domain has on average only 9% identity with the other three PAH domains (see Table 5.4A in the Appendix). This region is not identified as a PAH domain by domain detection programs such as NCBI's Conservative Domain search (Marcler-Bauer and Bryant, 2004) and different domain boundaries have been reported (Ayer et al., 1995; Halleck et al., 1995). The alignment of the PAH4 domain with the other PAH domain used here was manually optimized to match the structurally important residues of the PAH2 domain. However, several buried, hydrophobic residues, that are structurally important and highly conserved in PAH1-3, are replaced by hydrophilic and/or non-conserved residues in PAH4 (e.g. residues V209, L219 and L220 in PAH2; see Table 5.5A in the Appendix). Alternative alignments result in mismatches at other positions. This indicates that the PAH4 region most likely does not fold as a four helix bundle, but instead adapts a distinct fold. Contrastingly, the PAH1, PAH2 and PAH3 domains are highly conserved and have significant homology (cf. Table 5.4A). Residues in the core and in the interaction surface are conserved or conservatively mutated throughout these PAH domains (cf. Table 5.5A), indicating that the general fold and mode of interaction with the SID is conserved. Residues at the edges of the binding pocket (V152-I158, Q186 and E202) are less conserved, indicating that these residues play a role in defining the specificity of each member of the PAH domain family (cf. Table 5.5A and Figure 5.9A in the Appendix). Notably, A157 of $PAH2^B$ that is part of the interaction surface, is conserved in the PAH1 domain but is replaced by a conserved glutamic acid residue in PAH3, suggesting a role for electrostatic interaction with the SID. Furthermore, the bulky and buried aromatic side chains in the helix $\alpha1,\alpha2$ interface (F178 and Y160) and the buried side chains in the $\alpha3,\alpha4$ interface (V205, V209, and F223) are conserved between the different PAH2 sequences and conserved or conservatively mutated throughout the PAH1 and

PAH3 domains (cf. Table 5.5A). This strongly suggests that the ‘open’ conformation is an intrinsic feature of the PAH domain. The solution structure of the PAH1 domain of mSin3B bound to the SID of the NRSF/REST repressor, published after submission of our manuscript, shows that the complex is very similar to the structures of the PAH2 domain complexes: the PAH1 domain is folded as a four helix bundle and the SID is wedged between helices α1 and α2 (Nomura et al., 2005). NMR spectra of the unbound PAH1 domain both in this study and in our own experiments (data not shown) are of similar quality as for the extended PAH2^B construct, indicating that the unbound PAH1 domain is probably folded as well. Thus, on the basis of our structure of the PAH2^B domain and the multiple sequence alignment of the PAH domains, we suggest that the PAH1, PAH2 and PAH3 domains can form folded modules in full length Sin3, like ‘beads-on-a-string’, and that the PAH4 domain probably has a distinct fold.

Conclusion

The solution structure of the unbound PAH2 domain shows that the domain is folded and that its conformation does not change appreciably upon binding of its interaction partners, either the Mad1-SID or HBP1-SID. Notably, the binding pocket is accessible in the unbound structure as a result of conserved bulky residues in the helix α1,α2 interface that prevent close packing of α1 with the core. In addition, the orientation of crucial interacting side chains is highly similar in bound and unbound state, suggesting that they guide folding of the unstructured SID by defining a folding template. A more pronounced change is seen in the dynamics of unbound PAH2. The N-terminal part of helix α1 is rigidified by intermolecular interactions with the SID. Furthermore, conformational exchange with a minor, partially unfolded state is quenched when the SID is bound, indicating a stabilization of the folded state of the domain. On the basis of our data and multiple sequence alignment of the PAH domains, we suggest that the PAH1, PAH2 and PAH3 domains form folded binding modules in full length Sin3 like a ‘beads-on-a-string’ and act as a folding template for the SID in complex formation.

Materials and Methods

Sample Preparation

All NMR studies were performed using a protein construct of 85 or 105 residues (denoted as PAH2^B) corresponding to residues 148-232 (PAH2^B-85) and 148-252 (PAH2^B-105), respectively, of the long variant of Mm. mSin3B (SpTrEMBL accession number Q62141). Cloning, expression and purification of the PAH2^B domain were performed as described before (Spronk et al., 2001). Uniformly ^{15}N/^{13}C-labelled PAH2^B was prepared using [^{15}N]-NH_4Cl and [$^{13}C_6$]-glucose as sole nitrogen and carbon sources. NMR samples contained 0.5 to 0.9 mM ^{13}C/^{15}N double-labeled protein in a buffer of 50 mM KH_2PO_4 / K_2HPO_4 (pH 6.3), 100 mM KCl, H_2O / ^2H_2O (95% / 5%), using protease inhibitor Pefabloc and trace NaN_3 as preservative.

NMR spectroscopy

All NMR experiments were carried out on Varian Unity Inova 500, 600 and 800 MHz spectrometers. The data were processed using the NMRPipe suite (Delaglio et al., 1995) and analyzed using XEASY (Bartels et al., 1995) and Sparky (Goddard, T.D. & Kneller, D. G. (2004). SPARKY 3, University of California, San Francisco, http://www.cgl.ucsf.edu/home/sparky/). Assignments were obtained using 3D HNCA, HNCACB, CBCACONH, HNCAHA, (H)CCH-TOCSY, CBCDHD, ^{15}N NOESY-HSQC and aliphatic and aromatic ^{13}C NOESY-HSQC spectra. Stereo-specific assignments of the pro-*R* and pro-*S* methyl groups of Val and Leu-residues were made based on a CT-^{13}C-HSQC of 10% ^{13}C labeled PAH2^B-sample (Neri et al., 1989). Distance restraints for structure calculations were obtained from the 3D ^{15}N- and ^{13}C-separated NOESY spectra. The NOE mixing time in all NOESY experiments was set to 100 ms. Backbone N, HN, C_α, C_β, C’, and H_α atoms of unfolded PAH2 were assigned using a HNCACB, HNCO and HNCAHA experiments.
The ^{15}N-T_1, ^{15}N-$T_{1\rho}$ and {^{1}H}-^{15}N NOE experiments were recorded at 20.0 °C at 14.1 T. All experiments were recorded in an interleaved manner. Relaxation delays were: 16, 256 (2x), 384, 512, 768 (2x) and 1024 ms for the T_1-experiment and 16, 32 (2x), 48, 64, 96 (2x) and 128 ms for the $T_{1\rho}$-experiment. For both experiment the recycle delay was 1.55 s. The {^{1}H}-^{15}N

NOE values were derived from experiments using 3 s either on- or off-resonance irradiation in a total recycle delay of 5 sec. Additionally, a 12 s delay was executed after acquisition of each complex point. Relaxation dispersion experiments were recorded at 20 °C for ^{15}N-SQ coherence at 11.7 (duplicate experiments), 14.1 (duplicate experiments) and 18.8 T (triplicate experiments) using a standard constant time relaxation compensated CPMG scheme (Loria et al., 1999; Mulder et al., 2001). The constant-time value was set to 40 ms and the effective CPMG field strength was set to 50, 100, 200, 300, 400, 500, 700, 1000 and 1400 Hz using a ^{15}N 90° pulse of ca. 52 µs. The recycle delay was 2 s.

Structure calculations and refinement

The NOE peaks were assigned and converted into distance restraints by CYANA v1.05 (Güntert et al., 1997; Herrmann et al., 2002). Dihedral angle restraints for the ϕ-ψ angles were derived from N, HN, C', C_α, C_β, and H_α chemical shifts using TALOS (Cornilescu et al., 1999). An initial set of 100 NMR structures was calculated with CYANA and the 70 best structures were subsequently refined in water using a restrained molecular dynamics protocol (Spronk et al., 2002; Linge et al., 2003) to improve local geometry and electrostatics. Of the resulting structures, the 30 lowest energy structures with no distance restraint violations > 0.5 Å and dihedral angle violations > 5° were selected to form the final ensemble. The best representative structure (no. 1 of the ensemble) is defined as the structure closest to the mean structure. Structures were analyzed using the programs PROCHECK-NMR (Laskowski et al., 1996) and WHAT-CHECK (Hooft et al., 1996).

Fast time scale dynamics analysis

All spectra were processed and zero filled to a final size of 8K (^{1}H) x 4K (^{15}N) points. The longitudinal and transverse relaxation rates were calculated by fitting peak intensities to a mono-exponential decay using the modelXY module of NMRPipe (Delaglio et al., 1995). Errors in peak heights were estimated by the root-mean-square of the noise-level. Errors (1s confidence limits) were estimated from 21 Monte Carlo simulations. In a conservative approach, all errors for R_1, $R_{1\rho}$ and NOE were estimated to be at least 3% to prevent falsely attributed contributions of internal motions or exchange. Subsequently, the R_1, $R_{1\rho}$ and NOE values were analyzed using reduced spectral density mapping to describe the motion of the N-H vector in terms of the spectral density at 0, ω_N and $0.87\omega_H$ frequencies (Farrow et al, 1995; Lefèvre et al., 1996; Ishima and Nagayama et al., 1996). The chemical shift anisotropy of the ^{15}N nucleus was assumed uniform with a value of -170 ppm and the N-H bond length was set to 1.02 Å.

Slow time scale dynamics analysis

All spectra were processed and zero filled to a final size of 8K (^{1}H) x 4K (^{15}N) points. The effective R_2 was calculated based on peak intensities (Mulder et al., 2001). Uncertainties in peak intensities were based on the root-mean-square of the noise-level. The dispersion of R_2 values was defined as significant when the RMSD of a fit to a straight line was larger than twice the error based on noise. The program GLOVE (by J.C. Lansing and P.E. Wright, The Scripps Research Institute, La Jolla, CA, USA) was used to fit the effective R_2 values for all data sets simultaneous to the Carver-Richards-Jen all time scale approximation (Carver and Richards, 1972; Jen, 1978). First, all residues were fitted individually and the reduced chi-squared value of residue i, $\chi^2_{i,individual}$, was calculated. Then, residues in the same secondary structure element were fitted simultaneous in a global fit and the reduced chi-squared value of residue i in the global fit, $\chi^2_{i,global}$, was calculated. In the global fit the exchange rate constant kex and the populations p_a and p_b were optimized globally while the chemical shift difference $\Delta\omega$ and the exchange free R_2 were optimized for each fitted residue. If necessary, these elements were redefined until the global fit was accepted, defined as $\chi^2_{i,global} < 1.5 \cdot \chi^2_{i,individual}$. Subsequently, these elements were iteratively combined in global fits as long as these fits were accepted.

Databank Accession Numbers

The coordinates of the structures were deposited in the RCSB Protein Data Bank under PDB accession number 2F05. The chemical shifts and relaxation data of the free PAH2^B domain are deposited in the BioMagResBank under BMRB accession number 6899.

Appendix

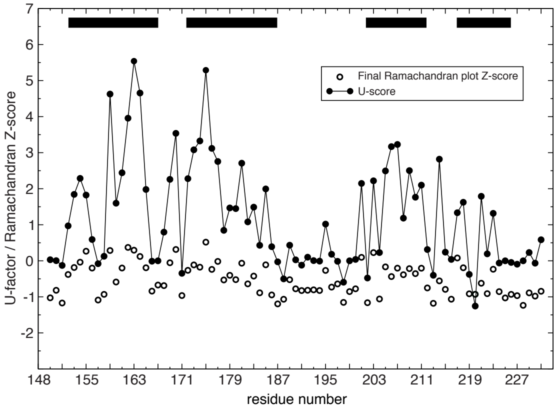


Figure 5.7A. Per-residue *U*-factor and Ramachandran quality *Z*-scores. Secondary structure is indicated by filled boxes (α-helix) at the top of the figure.

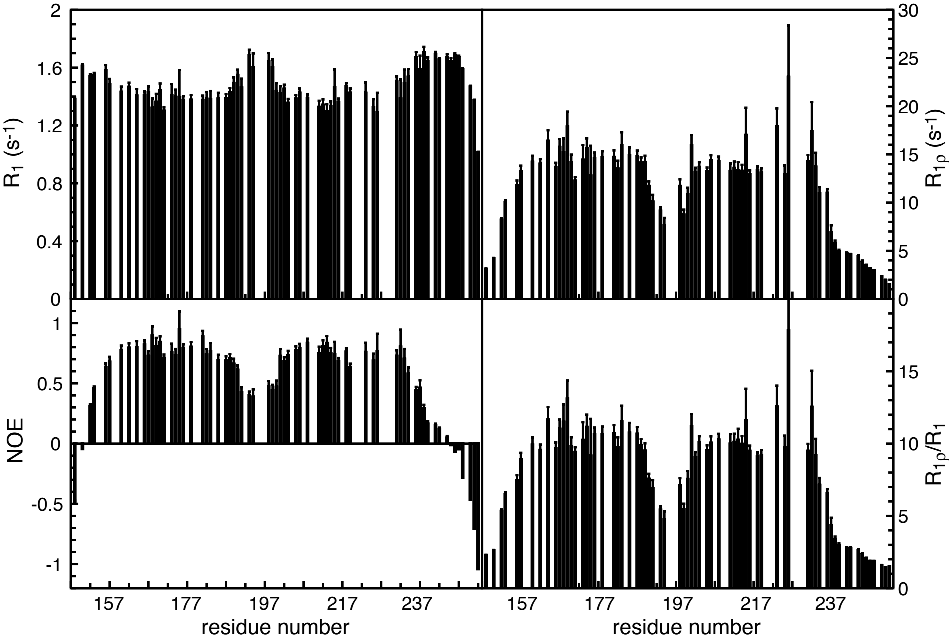


Figure 5.8A. ^{15}N relaxation parameters of the free PAH2^{B} domain recorded at 600 MHz proton frequency and 293K. R_1 (top left panel), $R_{1\rho}$ (top right panel), NOE (bottom left panel) and the $R_{1\rho}/R_1$ ratio (bottom right panel).

Table 5.3A.
Results from ^{15}N-R$_2$ relaxation dispersion data fitting[a].

res.	χ^2 [b]	$\Delta\omega$ (10^1 rad/s)[c]	error $\Delta\omega$ (10^1 rad/s)	k_{ex} (10^1 s^{-1})	error k_{ex} (10^1 s$^{-1)}$	p_b [d] (%)	error p_b (%)
152	0.83	51	2	130	5	0.92	0.05
153	0.66	62	3	130	5	0.92	0.05
157	0.75	61	3	130	5	0.92	0.05
160	0.60	62	4	130	5	0.92	0.05
162	0.51	56	3	130	5	0.92	0.05
163	0.91	56	5	130	5	0.92	0.05
167	0.95	83	5	130	5	0.92	0.05
168	0.70	116	7	229	13	1.20	0.07
169	0.39	184	13	229	13	1.20	0.07
170	1.15	99	5	229	13	1.20	0.07
171	0.85	33	2	229	13	1.20	0.07
173	0.53	86	8	130	5	0.92	0.05
174	0.60	82	6	130	5	0.92	0.05
176	0.93	40	3	130	5	0.92	0.05
178	0.82	56	3	130	5	0.92	0.05
181	0.76	46	3	130	5	0.92	0.05
183	0.53	62	7	130	5	0.92	0.05
185	0.60	40	4	130	5	0.92	0.05
187	0.68	45	3	130	5	0.92	0.05
188	0.67	47	3	130	5	0.92	0.05
189	0.53	66	4	130	5	0.92	0.05
190	0.66	59	3	130	5	0.92	0.05
191	1.10	73	5	130	5	0.92	0.05
193	1.13	61	3	130	5	0.92	0.05
198	0.59	116	8	130	5	0.92	0.05
200	0.57	57	4	130	5	0.92	0.05
201	0.75	130	10	130	5	0.92	0.05
202	0.56	32	3	130	5	0.92	0.05
203	0.79	42	3	130	5	0.92	0.05
206	0.52	55	3	130	5	0.92	0.05
208	0.89	45	3	130	5	0.92	0.05
213	0.44	50	4	130	5	0.92	0.05
214	0.39	36	3	130	5	0.92	0.05
215	0.42	140	23	130	5	0.92	0.05
216	1.01	33	2	130	5	0.92	0.05
218	0.70	37	2	130	5	0.92	0.05
219	1.14	84	4	130	5	0.92	0.05
223	0.56	210	17	229	13	1.20	0.07
226	0.35	173	28	229	13	1.20	0.07
231	0.68	124	6	229	13	1.20	0.07
232	0.31	196	22	229	13	1.20	0.07
234	0.95	73	4	229	13	1.20	0.07
236	1.31	85	4	229	13	1.20	0.07

[a] Global fits of the ^{15}N R$_2$-relaxation dispersion profiles to the Carver-Richards-Jen all-time scale approximation were performed as discussed in Materials and Methods.

[b] The reduced χ^2 for the global fit.

[c] The magnitude of the chemical shift difference, expressed in rad/s at 500 MHz proton frequency.

[d] population of the minor conformation.

Table 5.4A.
Sequence identity (%) between the PAH domains[a,b].

	PAH1	PAH2	PAH3	PAH4
PAH1	85 ± 9 / 52 ± 23	32 ± 1	22 ± 2	13 ± 2
PAH2	28 ± 8	78 ± 16 / 50 ± 22	19 ± 2	10 ± 2
PAH3	20 ± 6	20 ± 6	78 ± 16 / 34 ± 16	14 ± 2
PAH4	9 ± 3	9 ± 4	8 ± 3	67 ± 19 19 ± 16 /

[a] Top half compares the PAH domain in the vertebrates *Homo Sapiens*, *Mus Musculus* and *Xenopus Laevis*, bottom half compares all PAH domains (accession numbers in Figure 5.9A).

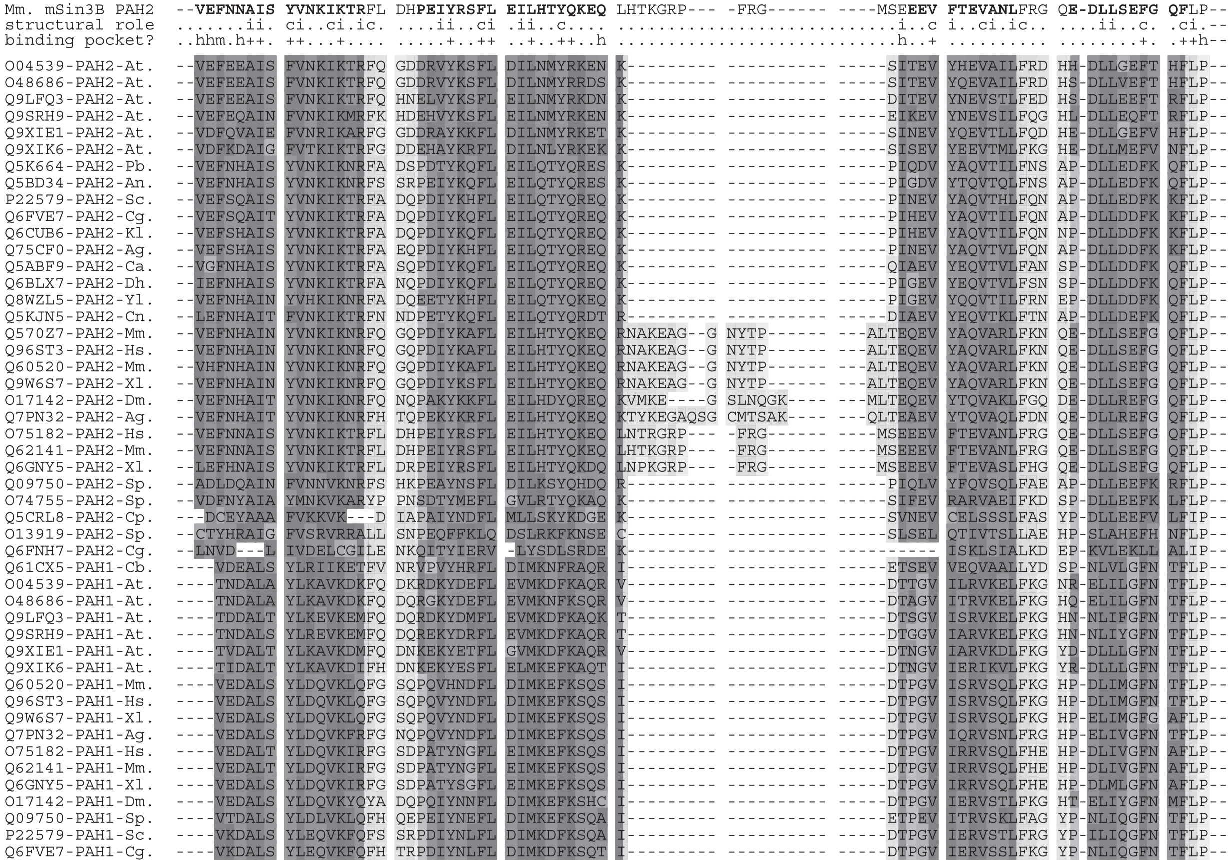
[b] Domain boundaries in Mm. mSin3B: 34-98 (PAH1), 152-228 (PAH2), 293-358 (PAH3), 705-789 (PAH4).

Table 5.5A.
Sequence diversity of residues in the binding pocket of the PAH domains[a].

residue	role[b]	PAH1	PAH2	PAH3	PAH4
F154	BP	v,t,c	f	t,l,y,(x)	v,i,a,(x)
A157	i12, BP	**A**	**A**	e,d	l,e,n
I158	i12, BP	**L**	**I**	x	f,e
Y160	i12/14, BP	**Y**	**Y,F**	**F**	x (a,e,c)
V161	i12, BP	**L**	**V**	F(L,C)	n, l
I164	i12/i14	**V,L**	**I,V**	**V(A)**	I,W(T,N,M)
K165	i12, BP	**K**	**K**	**K,R**	y,g
F168	core	**F**	**F**	L,I	F,E,L
I174	i23	V(K,I,T)	I(T,V,K)	V,A,T,I,D	L,V
Y175	i12,BP	**Y**	**Y**	**Y,F**	Y,C
F178	i23/i12, BP	**F**	**F**	**F**	L(N)
L179	i12,BP	**L**	**L**	**L**	**L(X)**
I181	i23	**I,V**	**I**	**C,L**	**I,V(K,E)**
L182	i12,BP	**M**	**L**	**L,I**	k,y
Y185	core, i23, BP	**F**	**Y**	**F,Y**	k,a,e,n
Q186	BP	K	Q,R	s,n(v)	x
E202	i34, BP	t	i,e	r,k(g,t)	d,x
V205	i23/34, BP	**V**	**V**	**L**	L,A,M(E,D)
F206	i34	I	Y,F	V,L,Q	I,X
V209	i23/34	**V**	**V**	**V,A**	r,q,x
A210	i34	S(K)	T,A,S	S(X)	l,x
L212	i23	**L**	**L**	F,Y,L	d,e,(x)
F213	core	**F**	**F**	L,I	M,L(X)
L219	i14	**L**	**L**	**L**	a,s,x
L220	i34	**I,L**	**L**	F,M,L(W)	e,s,x
F223	core, i34, BP	**F**	**F**	**F**	Y(E)
F226	core, BP	**F**	**F**	**F,L**	F,L,V(I,T)
L227	i34, BP	**L**	**L**	**L,V,M**	L,V(Y,K,E)

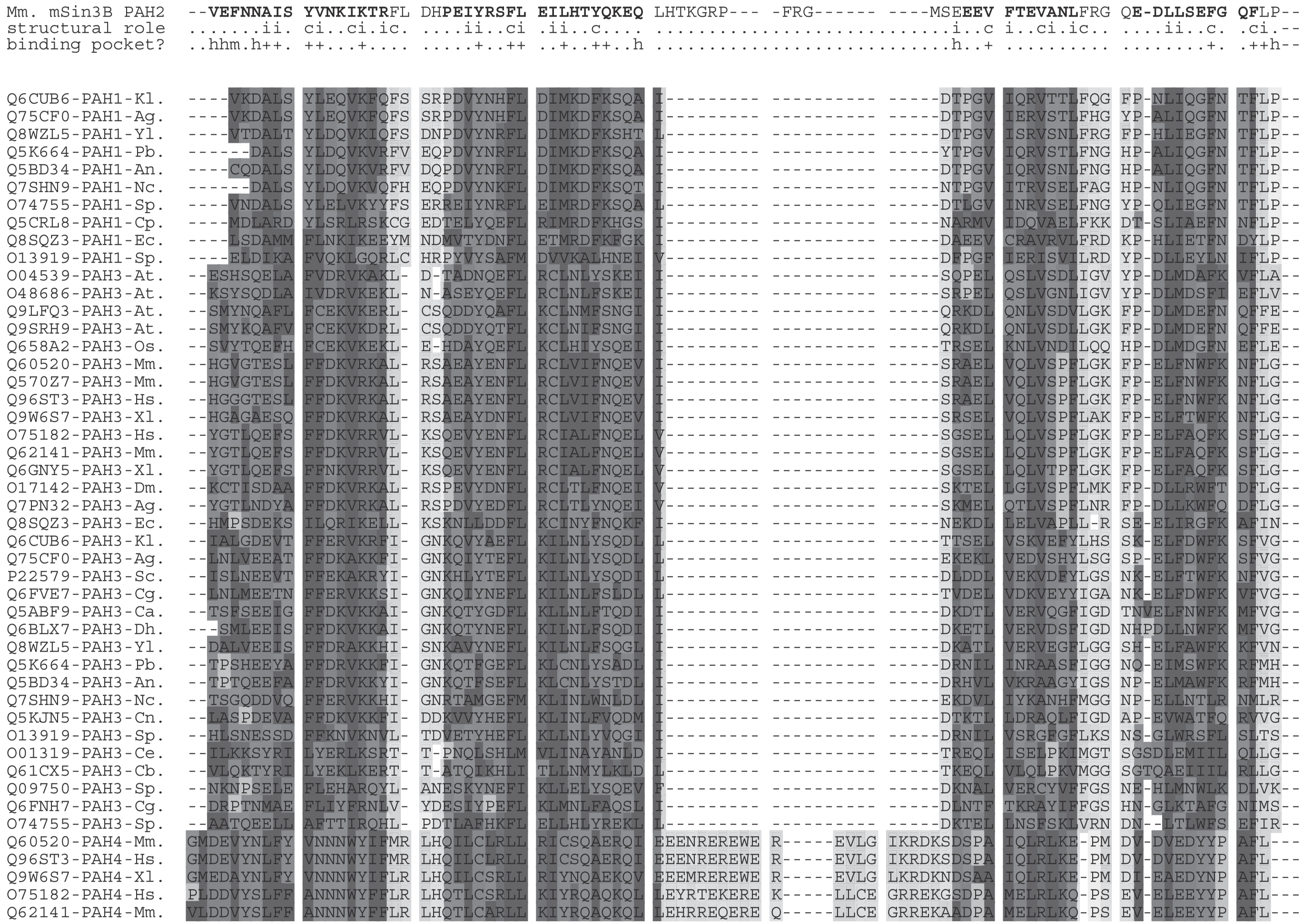
[a] Most frequent amino acid types are listed that occur at a certain position in the PAH*i* sequences (*i*=1,2,3,4) (80% cut-off). Residues are listed in order of decreasing occurrence. When some amino acids occur in the majority of sequences, the ones occurring in the minority are placed between round brackets. X denotes any amino acid. Similarities with the Mm. PAH2^{B} sequence are shown in uppercase, dissimilarities in italic lowercase. Strong similarities are shown in bold-type, while strong dissimilarities are shown with underline.

[b] BP = part of binding pocket; i12, i23, i34, i14 = part of interface between helix α1 and α2, α2 and α3, etc.; core = part of protein core.



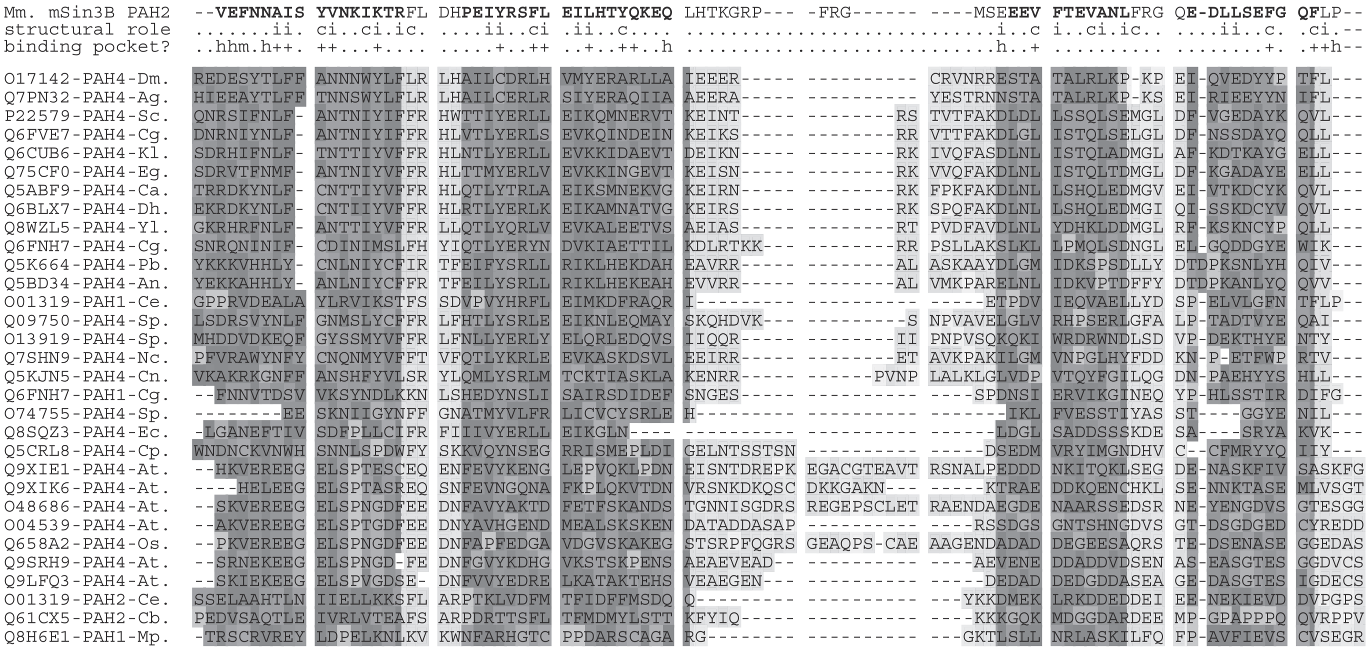
```
Mm. mSin3B PAH2   --VEFNNAIS YVNKIKTRFL DHPEIYRSFL EILHTYQKEQ LHTKGRP--- -FRG------ ----MSEEEV FTEVANLFRG QE-DLLSEFG QFLP--
structural role   .......ii. ci..ci.ic. ....ii..ci .ii..c.... .......... .......... ......i..c i..ci.ic.. ....ii..c. .ci.--
binding pocket?   ..hhm.h++. ++...+.... .....+..++ ..+..++..h .......... .......... ......h..+ .......... ........+. .++h--

O04539-PAH2-At.   --VEFEEAIS FVNKIKTRFQ GDDRVYKSFL DILNMYRKEN K--------- ---------- -----SITEV YHEVAILFRD HH-DLLGEFT HFLP--
O48686-PAH2-At.   --VEFEEAIS FVNKIKTRFQ GDDRVYKSFL DILNMYRRDS K--------- ---------- -----SITEV YQEVAILFRD HS-DLLVEFT HFLP--
Q9LFQ3-PAH2-At.   --VEFEEAIS FVNKIKTRFQ HNELVYKSFL EILNMYRKDN K--------- ---------- -----DITEV YNEVSTLFED HS-DLLEEFT RFLP--
Q9SRH9-PAH2-At.   --VEFEQAIN FVNKIKMRFK HDEHVYKSFL EILNMYRKEN K--------- ---------- -----EIKEV YNEVSILFQG HL-DLLEQFT RFLP--
Q9XIE1-PAH2-At.   --VDFQVAIE FVNRIKARFG GDDRAYKKFL DILNMYRKET K--------- ---------- -----SINEV YQEVTLLFQD HE-DLLGEFV HFLP--
Q9XIK6-PAH2-At.   --VDFKDAIG FVTKIKTRFG DDEHAYKRFL DILNLYRKEK K--------- ---------- -----SISEV YEEVTMLFKG HE-DLLMEFV NFLP--
Q5K664-PAH2-Pb.   --VEFNHAIS YVNKIKNRFA DSPDTYKQFL EILQTYQRES K--------- ---------- -----PIQDV YAQVTILFNS AP-DLLEDFK QFLP--
Q5BD34-PAH2-An.   --VEFNHAIS YVNKIKNRFS SRPEIYKQFL EILQTYQRES K--------- ---------- -----PIGDV YTQVTQLFNS AP-DLLEDFK QFLP--
P22579-PAH2-Sc.   --VEFSQAIS YVNKIKTRFA DQPDIYKHFL EILQTYQREQ K--------- ---------- -----PINEV YAQVTHLFQN AP-DLLEDFK KFLP--
Q6FVE7-PAH2-Cg.   --VEFSQAIT YVNKIKTRFA DQPDIYKQFL EILQTYQREQ K--------- ---------- -----PIHEV YAQVTILFQN AP-DLLDDFK KFLP--
Q6CUB6-PAH2-Kl.   --VEFSHAIS YVNKIKTRFA DQPDIYKQFL EILQTYQREQ K--------- ---------- -----PIHEV YAQVTILFQN NP-DLLDDFK KFLP--
Q75CF0-PAH2-Ag.   --VEFSHAIS YVNKIKTRFA EQPDIYKHFL EILQTYQREQ K--------- ---------- -----PINEV YAQVTVLFQN AP-DLLDDFK KFLP--
Q5ABF9-PAH2-Ca.   --VGFNHAIS YVNKIKTRFA SQPDIYKQFL EILQTYQREQ K--------- ---------- -----QIAEV YEQVTVLFAN SP-DLLDDFK QFLP--
Q6BLX7-PAH2-Dh.   --IEFNHAIS YVNKIKTRFA NQPDIYKQFL EILQTYQREQ K--------- ---------- -----PIGEV YEQVTVLFSN SP-DLLDDFK QFLP--
Q8WZL5-PAH2-Yl.   --VEFNHAIN YVHKIKNRFA DQEETYKHFL EILQTYQKEQ K--------- ---------- -----PIGEV YQQVTILFRN EP-DLLDDFK QFLP--
Q5KJN5-PAH2-Cn.   --LEFNHAIT FVNKIKTRFN NDPETYKQFL EILQTYQRDT R--------- ---------- -----DIAEV YEQVTKLFTN AP-DLLDEFK QFLP--
Q570Z7-PAH2-Mm.   --VEFNHAIN YVNKIKNRFQ GQPDIYKAFL EILHTYQKEQ RNAKEAG--G NYTP------ ---ALTEQEV YAQVARLFKN QE-DLLSEFG QFLP--
Q96ST3-PAH2-Hs.   --VEFNHAIN YVNKIKNRFQ GQPDIYKAFL EILHTYQKEQ RNAKEAG--G NYTP------ ---ALTEQEV YAQVARLFKN QE-DLLSEFG QFLP--
Q60520-PAH2-Mm.   --VHFNHAIN YVNKIKNRFQ GQPDIYKAFL EILHTYQKEQ RNAKEAG--G NYTP------ ---ALTEQEV YAQVARLFKN QE-DLLSEFG QFLP--
Q9W6S7-PAH2-Xl.   --VEFNHAIN YVNKIKNRFQ GQPDIYKSFL EILHTYQKEQ RNAKEAG--G NYTP------ ---ALTEQEV YAQVARLFKN QE-DLLSEFG QFLP--
O17142-PAH2-Dm.   --VEFNHAIT YVNKIKNRFQ NQPAKYKKFL EILHDYQREQ KVMKE----G SLNQGK---- ---MLTEQEV YTQVAKLFGQ DE-DLLREFG QFLP--
Q7PN32-PAH2-Ag.   --VEFNHAIT YVNKIKNRFH TQPEKYKRFL EILHTYQKEQ KTYKEGAQSG CMTSAK---- ---QLTEAEV YTQVAQLFDN QE-DLLREFG QFLP--
O75182-PAH2-Hs.   --VEFNNAIS YVNKIKTRFL DHPEIYRSFL EILHTYQKEQ LNTRGRP--- -FRG------ ----MSEEEV FTEVANLFRG QE-DLLSEFG QFLP--
Q62141-PAH2-Mm.   --VEFNNAIS YVNKIKTRFL DHPEIYRSFL EILHTYQKEQ LHTKGRP--- -FRG------ ----MSEEEV FTEVANLFRG QE-DLLSEFG QFLP--
Q6GNY5-PAH2-Xl.   --LEFHNAIS YVNKIKTRFL DRPEIYRSFL EILHTYQKDQ LNPKGRP--- -FRG------ ----MSEEEV FTEVASLFHG QE-DLLSEFG QFLP--
Q09750-PAH2-Sp.   --ADLDQAIN FVNNVKNRFS HKPEAYNSFL DILKSYQHDQ R--------- ---------- -----PIQLV YFQVSQLFAE AP-DLLEEFK RFLP--
O74755-PAH2-Sp.   --VDFNYAIA YMNKVKARYP PNSDTYMEFL GVLRTYQKAQ K--------- ---------- -----SIFEV RARVAEIFKD SP-DLLEEFK LFLP--
Q5CRL8-PAH2-Cp.   ---DCEYAAA FVKKVK---D IAPAIYNDFL MLLSKYKDGE K--------- ---------- -----SVNEV CELSSSLFAS YP-DLLEEFV LFIP--
O13919-PAH2-Sp.   --CTYHRAIG FVSRVRRALL SNPEQFFKLQ DSLRKFKNSE C--------- ---------- -----SLSEL QTIVTSLLAE HP-SLAHEFH NFLP--
Q6FNH7-PAH2-Cg.   --LNVD---L IVDELCGILE NKQITYIERV -LYSDLSRDE K--------- ---------- ---------- ISKLSIALKD EP-KVLEKLL ALIP--
Q61CX5-PAH1-Cb.   ----VDEALS YLRIIKETFV NRVPVYHRFL DIMKNFRAQR I--------- ---------- -----ETSEV VEQVAALLYD SP-NLVLGFN TFLP--
O04539-PAH1-At.   ----TNDALA YLKAVKDKFQ DKRDKYDEFL EVMKDFKAQR V--------- ---------- -----DTTGV ILRVKELFKG NR-ELILGFN TFLP--
O48686-PAH1-At.   ----TNDALA YLKAVKDKFQ DQRGKYDEFL EVMKNFKSQR V--------- ---------- -----DTAGV ITRVKELFKG HQ-ELILGFN TFLP--
Q9LFQ3-PAH1-At.   ----TDDALT YLKEVKEMFQ DQRDKYDMFL EVMKDFKAQK T--------- ---------- -----DTSGV ISRVKELFKG HN-NLIFGFN TFLP--
Q9SRH9-PAH1-At.   ----TNDALS YLREVKEMFQ DQREKYDRFL EVMKDFKAQR T--------- ---------- -----DTGGV IARVKELFKG HN-NLIYGFN TFLP--
Q9XIE1-PAH1-At.   ----TVDALT YLKAVKDMFQ DNKEKYETFL GVMKDFKAQR V--------- ---------- -----DTNGV IARVKDLFKG YD-DLLLGFN TFLP--
Q9XIK6-PAH1-At.   ----TIDALT YLKAVKDIFH DNKEKYESFL ELMKEFKAQT I--------- ---------- -----DTNGV IERIKVLFKG YR-DLLLGFN TFLP--
Q60520-PAH1-Mm.   ----VEDALS YLDQVKLQFG SQPQVHNDFL DIMKEFKSQS I--------- ---------- -----DTPGV ISRVSQLFKG HP-DLIMGFN TFLP--
Q96ST3-PAH1-Hs.   ----VEDALS YLDQVKLQFG SQPQVYNDFL DIMKEFKSQS I--------- ---------- -----DTPGV ISRVSQLFKG HP-DLIMGFN TFLP--
Q9W6S7-PAH1-Xl.   ----VEDALS YLDQVKLQFG SQPQVYNDFL DIMKEFKSQS I--------- ---------- -----DTPGV ISRVSQLFKG HP-ELIMGFG AFLP--
Q7PN32-PAH1-Ag.   ----VEDALS YLDQVKFRFG NQPQVYNDFL DIMKEFKSQS I--------- ---------- -----DTPGV IQRVSNLFRG HP-ELIVGFN TFLP--
O75182-PAH1-Hs.   ----VEDALT YLDQVKIRFG SDPATYNGFL EIMKEFKSQS I--------- ---------- -----DTPGV IRRVSQLFHE HP-DLIVGFN AFLP--
Q62141-PAH1-Mm.   ----VEDALT YLDQVKIRFG SDPATYNGFL EIMKEFKSQS I--------- ---------- -----DTPGV IRRVSQLFHE HP-DLIVGFN AFLP--
Q6GNY5-PAH1-Xl.   ----VEDALS YLDQVKIRFG SDPATYSGFL EIMKEFKSQS I--------- ---------- -----DTPGV IRRVSQLFHE HP-DLMLGFN AFLP--
O17142-PAH1-Dm.   ----VEDALS YLDQVKYQYA DQPQIYNNFL DIMKEFKSHC I--------- ---------- -----DTPGV IERVSTLFKG HT-ELIYGFN MFLP--
Q09750-PAH1-Sp.   ----VTDALS YLDLVKLQFH QEPEIYNEFL DIMKEFKSQA I--------- ---------- -----ETPEV ITRVSKLFAG YP-NLIQGFN TFLP--
P22579-PAH1-Sc.   ----VKDALS YLEQVKFQFS SRPDIYNLFL DIMKDFKSQA I--------- ---------- -----DTPGV IERVSTLFRG YP-ILIQGFN TFLP--
Q6FVE7-PAH1-Cg.   ----VKDALS YLEQVKFQFH TRPDIYNLFL DIMKDFKSQT I--------- ---------- -----DTPGV IERVSSLFKG YP-NLIQGFN TFLP--
```



```
Mm. mSin3B PAH2    --VEFNNAIS YVNKIKTRFL DHPEIYRSFL EILHTYQKEQ LHTKGRP--- -FRG------ ----MSEEEV FTEVANLFRG QE-DLLSEFG QFLP--
structural role    .......ii. ci..ci.ic. ....ii..ci .ii..c.... ............................i..c i..ci.ic.. ....ii..c. .ci.--
binding pocket?    ..hhm.h++. ++...+.... .....+..++ ..+..++..h ............................h..+ .......... ........+. .++h--

Q6CUB6-PAH1-Kl.    ----VKDALS YLEQVKFQFS SRPDVYNHFL DIMKDFKSQA I--------- ---------- -----DTPGV IQRVTTLFQG FP-NLIQGFN TFLP--
Q75CF0-PAH1-Ag.    ----VKDALS YLEQVKFQFS SRPDVYNHFL DIMKDFKSQA I--------- ---------- -----DTPGV IERVSTLFHG YP-ALIQGFN TFLP--
Q8WZL5-PAH1-Yl.    ----VTDALT YLDQVKIQFS DNPDVYNRFL DIMKDFKSHT L--------- ---------- -----DTPGV ISRVSNLFRG FP-HLIEGFN TFLP--
Q5K664-PAH1-Pb.    ------DALS YLDQVKVRFV EQPDVYNRFL DIMKDFKSQA I--------- ---------- -----YTPGV IQRVSTLFNG HP-ALIQGFN TFLP--
Q5BD34-PAH1-An.    ----CQDALS YLDQVKVRFV DQPDVYNRFL DIMKDFKSQA I--------- ---------- -----DTPGV IQRVSNLFNG HP-ALIQGFN TFLP--
Q7SHN9-PAH1-Nc.    ------DALS YLDQVKVQFH EQPDVYNKFL DIMKDFKSQT I--------- ---------- -----NTPGV ITRVSELFAG HP-NLIQGFN TFLP--
O74755-PAH1-Sp.    ----VNDALS YLELVKYYFS ERREIYNRFL EIMRDFKSQA L--------- ---------- -----DTLGV INRVSELFNG YP-QLIEGFN TFLP--
Q5CRL8-PAH1-Cp.    ----MDLARD YLSRLRSKCG EDTELYQEFL RIMRDFKHGS I--------- ---------- -----NARMV IDQVAELFKK DT-SLIAEFN NFLP--
Q8SQZ3-PAH1-Ec.    ----LSDAMM FLNKIKEEYM NDMVTYDNFL ETMRDFKFGK I--------- ---------- -----DAEEV CRAVRVLFRD KP-HLIETFN DYLP--
O13919-PAH1-Sp.    ----ELDIKA FVQKLGQRLC HRPYVYSAFM DVVKALHNEI V--------- ---------- -----DFPGF IERISVILRD YP-DLLEYLN IFLP--
O04539-PAH3-At.    --ESHSQELA FVDRVKAKL- D-TADNQEFL RCLNLYSKEI I--------- ---------- -----SQPEL QSLVSDLIGV YP-DLMDAFK VFLA--
O48686-PAH3-At.    --KSYSQDLA IVDRVKEKL- N-ASEYQEFL RCLNLFSKEI I--------- ---------- -----SRPEL QSLVGNLIGV YP-DLMDSFI EFLV--
Q9LFQ3-PAH3-At.    --SMYNQAFL FCEKVKERL- CSQDDYQAFL KCLNMFSNGI I--------- ---------- -----QRKDL QNLVSDVLGK FP-DLMDEFN QFFE--
Q9SRH9-PAH3-At.    --SMYKQAFV FCEKVKDRL- CSQDDYQTFL KCLNIFSNGI I--------- ---------- -----QRKDL QNLVSDLLGK FP-DLMDEFN QFFE--
Q658A2-PAH3-Os.    --SVYTQEFH FCEKVKEKL- E-HDAYQEFL KCLHIYSQEI I--------- ---------- -----TRSEL KNLVNDILQQ HP-DLMDGFN EFLE--
Q60520-PAH3-Mm.    --HGVGTESL FFDKVRKAL- RSAEAYENFL RCLVIFNQEV I--------- ---------- -----SRAEL VQLVSPFLGK FP-ELFNWFK NFLG--
Q570Z7-PAH3-Mm.    --HGVGTESL FFDKVRKAL- RSAEAYENFL RCLVIFNQEV I--------- ---------- -----SRAEL VQLVSPFLGK FP-ELFNWFK NFLG--
Q96ST3-PAH3-Hs.    --HGGGTESL FFDKVRKAL- RSAEAYENFL RCLVIFNQEV I--------- ---------- -----SRAEL VQLVSPFLGK FP-ELFNWFK NFLG--
Q9W6S7-PAH3-Xl.    --HGAGAESQ FFDKVRKAL- RSAEAYDNFL RCLVIFNQEV I--------- ---------- -----SRSEL VQLVSPFLAK FP-ELFTWFK NFLG--
O75182-PAH3-Hs.    --YGTLQEFS FFDKVRRVL- KSQEVYENFL RCIALFNQEL V--------- ---------- -----SGSEL LQLVSPFLGK FP-ELFAQFK SFLG--
Q62141-PAH3-Mm.    --YGTLQEFS FFDKVRRVL- KSQEVYENFL RCIALFNQEL V--------- ---------- -----SGSEL LQLVSPFLGK FP-ELFAQFK SFLG--
Q6GNY5-PAH3-Xl.    --YGTLQEFS FFNKVRRVL- KSQEVYENFL RCIALFNQEL V--------- ---------- -----SGSEL LQLVTPFLGK FP-ELFAQFK SFLG--
O17142-PAH3-Dm.    --KCTISDAA FFDKVRKAL- RSPEVYDNFL RCLTLFNQEI V--------- ---------- -----SKTEL LGLVSPFLMK FP-DLLRWFT DFLG--
Q7PN32-PAH3-Ag.    --YGTLNDYA FFDKVRKAL- RSPDVYEDFL RCLTLYNQEI V--------- ---------- -----SKMEL QTLVSPFLNR FP-DLLKWFQ DFLG--
Q8SQZ3-PAH3-Ec.    --HMPSDEKS ILQRIKELL- KSKNLLDDFL KCINYFNQKF I--------- ---------- -----NEKDL LELVAPLL-R SE-ELIRGFK AFIN--
Q6CUB6-PAH3-Kl.    --IALGDEVT FFERVKRFI- GNKQVYAEFL KILNLYSQDL L--------- ---------- -----TTSEL VSKVEFYLHS SK-ELFDWFK SFVG--
Q75CF0-PAH3-Ag.    --LNLVEEAT FFDKAKKFI- GNKQIYTEFL KILNLYSQDL L--------- ---------- -----EKEKL VEDVSHYLSG SP-ELFDWFK SFVG--
P22579-PAH3-Sc.    --ISLNEEVT FFEKAKRYI- GNKHLYTEFL KILNLYSQDI L--------- ---------- -----DLDDL VEKVDFYLGS NK-ELFTWFK NFVG--
Q6FVE7-PAH3-Cg.    --LNLMEETN FFERVKKSI- GNKQIYNEFL KILNLFSLDL L--------- ---------- -----TVDEL VDKVEYYIGA NK-ELFDWFK VFVG--
Q5ABF9-PAH3-Ca.    --TSFSEEIG FFDKVKKAI- GNKQTYGDFL KLLNLFTQDI I--------- ---------- -----DKDTL VERVQGFIGD TNVELFNWFK MFVG--
Q6BLX7-PAH3-Dh.    ---SMLEEIS FFDKVKKAI- GNKQTYNEFL KILNLFSQDI I--------- ---------- -----DKETL VERVDSFIGD NHPDLLNWFK MFVG--
Q8WZL5-PAH3-Yl.    --DALVEEIS FFDRAKKHI- SNKAVYNEFL KILNLFSQGL I--------- ---------- -----DKATL VERVEGFLGG SH-ELFAWFK KFVN--
Q5K664-PAH3-Pb.    --TPSHEEYA FFDRVKKFI- GNKQTFGEFL KLCNLYSADL I--------- ---------- -----DRNIL INRAASFIGG NQ-EIMSWFK RFMH--
Q5BD34-PAH3-An.    --TPTQEEFA FFDRVKKYI- GNKQTFSEFL KLCNLYSTDL I--------- ---------- -----DRHVL VKRAAGYIGS NP-ELMAWFK RFMH--
Q7SHN9-PAH3-Nc.    --TSGQDDVQ FFERVKKHI- GNRTAMGEFM KLINLWNLDL I--------- ---------- -----EKDVL IYKANHFMGG NP-ELLNTLR NMLR--
Q5KJN5-PAH3-Cn.    --LASPDEVA FFDKVKKFI- DDKVVYHEFL KLINLFVQDM I--------- ---------- -----DTKTL LDRAQLFIGD AP-EVWATFQ RVVG--
O13919-PAH3-Sp.    --HLSNESSD FFKNVKNVL- TDVETYHEFL KLLNLYVQGI I--------- ---------- -----DRNIL VSRGFGFLKS NS-GLWRSFL SLTS--
O01319-PAH3-Ce.    --ILKKSYRI LYERLKSRT- T-PNQLSHLM VLINAYANLD I--------- ---------- -----TREQL ISELPKIMGT SGSDLEMIIL QLLG--
Q61CX5-PAH3-Cb.    --VLQKTYRI LYEKLKERT- T-ATQIKHLI TLLNMYLKLD L--------- ---------- -----TKEQL VLQLPKVMGG SGTQAEIIIL RLLG--
Q09750-PAH3-Sp.    --NKNPSELE FLEHARQYL- ANESKYNEFI KLLELYSQEV F--------- ---------- -----DKNAL VERCYVFFGS NE-HLMNWLK DLVK--
Q6FNH7-PAH3-Cg.    --DRPTNMAE FLIYFRNLV- YDESIYPEFL KLMNLFAQSL I--------- ---------- -----DLNTF TKRAYIFFGS HN-GLKTAFG NIMS--
O74755-PAH3-Sp.    --AATQEELL AFTTIRQHL- PDTLAFHKFL ELLHLYREKL L--------- ---------- -----DKTEL LNSFSKLVRN DN--LTLWFS EFIR--
Q60520-PAH4-Mm.    GMDEVYNLFY VNNNWYIFMR LHQILCLRLL RICSQAERQI EEENREREWE R-----EVLG IKRDKSDSPA IQLRLKE-PM DV-DVEDYYP AFL---
Q96ST3-PAH4-Hs.    GMDEVYNLFY VNNNWYIFMR LHQILCLRLL RICSQAERQI EEENREREWE R-----EVLG IKRDKSDSPA IQLRLKE-PM DV-DVEDYYP AFL---
Q9W6S7-PAH4-Xl.    GMEDAYNLFY VNNNWYIFLR LHQILCSRLL RIYNQAEKQV EEEMREREWE R-----EVLG LKRDKNDSAA IQLRLKE-PM DI-EAEDYNP AFL---
O75182-PAH4-Hs.    PLDDVYSLFF ANNNWYFFLR LHQTLCSRLL KIYRQAQKQL LEYRTEKERE K-----LLCE GRREKGSDPA MELRLKQ-PS EV-ELEEYYP AFL---
Q62141-PAH4-Mm.    VLDDVYSLFF ANNNWYFFLR LHQTLCARLL KIYRQAQKQL LEHRREQERE Q-----LLCE GRREKAADPA MELRLKQ-PS EV-ELEEYYP AFL---
```



```
Mm. mSin3B PAH2   --VEFNNAIS YVNKIKTRFL DHPEIYRSFL EILHTYQKEQ LHTKGRP--- -FRG------ ----MSEEEV FTEVANLFRG QE-DLLSEFG QFLP--
structural role   .......ii. ci..ci.ic. ....ii..ci .ii..c.... .......... .......... ......i..c i..ci.ic.. ....ii..c. .ci.--
binding pocket?   ..hhm.h++. ++...+.... .....+..++ ..+..++..h .......... .......... ......h..+ .......... ........+. .++h--

O17142-PAH4-Dm.   REDESYTLFF ANNNWYLFLR LHAILCDRLH VMYERARLLA IEEER----- ---------- CRVNRRESTA TALRLKP-KP EI-QVEDYYP TFL---
Q7PN32-PAH4-Ag.   HIEEAYTLFF TNNSWYLFLR LHAILCERLR SIYERAQIIA AEERA----- ---------- YESTRNNSTA TALRLKP-KS EI-RIEEYYN IFL---
P22579-PAH4-Sc.   QNRSIFNLF- ANTNIYIFFR HWTTIYERLL EIKQMNERVT KEINT----- --------RS TVTFAKDLDL LSSQLSEMGL DF-VGEDAYK QVL---
Q6FVE7-PAH4-Cg.   DNRNIYNLF- ANTNIYIFFR HLVTLYERLS EVKQINDEIN KEIKS----- --------RR VTTFAKDLGL ISTQLSELGL DF-NSSDAYQ QLL---
Q6CUB6-PAH4-Kl.   SDRHIFNLF- TNTTIYVFFR HLNTLYERLL EVKKIDAEVT DEIKN----- --------RK IVQFASDLNL ISTQLADMGL AF-KDTKAYG ELL---
Q75CF0-PAH4-Eg.   SDRVTFNMF- ANTNIYVFFR HLTTMYERLV EVKKINGEVT KEISN----- --------RK VVQFAKDLNL ISTQLTDMGL DF-KGADAYE ELL---
Q5ABF9-PAH4-Ca.   TRRDKYNLF- CNTTIYVFFR HLQTLYTRLA EIKSMNEKVG KEIRN----- --------RK FPKFAKDLNL LSHQLEDMGV EI-VTKDCYK QVL---
Q6BLX7-PAH4-Dh.   EKRDKYNLF- CNTIIYVFFR HLRTLYERLK EIKAMNATVG KEIRS----- --------RK SPQFAKDLNL LSHQLEDMGI QI-SSKDCYV QVL---
Q8WZL5-PAH4-Yl.   GKRHRFNLF- ANTTIYVFFR LLQTLYQRLV EVKALEETVS AEIAS----- --------RT PVDFAVDLNL YDHKLDDMGL RF-KSKNCYP QLL---
Q6FNH7-PAH4-Cg.   SNRQNINIF- CDINIMSLFH YIQTLYERYN DVKIAETTIL KDLRTKK--- --------RR PSLLAKSLKL LPMQLSDNGL EL-GQDDGYE WIK---
Q5K664-PAH4-Pb.   YKKKVHHLY- CNLNIYCFIR TFEIFYSRLL RIKLHEKDAH EAVRR----- --------AL ASKAAYDLGM IDKSPSDLLY DTDPKSNLYH QIV---
Q5BD34-PAH4-An.   YEKKAHHLY- ANLNIYCFFR TFEILYSRLL RIKLHEKEAH EVVRR----- --------AL VMKPARELNL IDKVPTDFFY DTDPKANLYQ QVV---
O01319-PAH1-Ce.   GPPRVDEALA YLRVIKSTFS SDVPVYHRFL EIMKDFRAQR I--------- ---------- -----ETPDV IEQVAELLYD SP-ELVLGFN TFLP--
Q09750-PAH4-Sp.   LSDRSVYNLF GNMSLYCFFR LFHTLYSRLE EIKNLEQMAY SKQHDVK--- ---------S NPVAVELGLV RHPSERLGFA LP-TADTVYE QAI---
O13919-PAH4-Sp.   MHDDVDKEQF GYSSMYVFFR LFNLLYERLY ELQRLEDQVS IIQQR----- --------II PNPVSQKQKI WRDRWNDLSD VP-DEKTHYE NTY---
Q7SHN9-PAH4-Nc.   PFVRAWYNFY CNQNMYVFFT VFQTLYKRLE EVKASKDSVL EEIRR----- --------ET AVKPAKILGM VNPGLHYFDD KN-P-ETFWP RTV---
Q5KJN5-PAH4-Cn.   VKAKRKGNFF ANSHFYVLSR YLQMLYSRLM TCKTIASKLA KENRR----- ------PVNP LALKLGLVDP VTQYFGILQG DN-PAEHYYS HLL---
Q6FNH7-PAH1-Cg.   --FNNVTDSV VKSYNDLKKN LSHEDYNSLI SAIRSDIDEF SNGES----- ---------- ----SPDNSI ERVIKGINEQ YP-HLSSTIR DIFG--
O74755-PAH4-Sp.   --------EE SKNIIGYNFF GNATMYVLFR LICVCYSRLE H--------- ---------- -------IKL FVESSTIYAS ST----GGYE NIL---
Q8SQZ3-PAH4-Ec.   -LGANEFTIV SDFPLLCIFR FIIIVYERLL EIKGLN---- ---------- ---------- ------LDGL SADDSSSKDE SA----SRYA KVK---
Q5CRL8-PAH4-Cp.   WNDNCKVNWH SNNLSPDWFY SKKVQYNSEG RRISMEPLDI GELNTSSTSN ---------- -----DSEDM VRYIMGNDHV C--CFMRYYQ IIY---
Q9XIE1-PAH4-At.   --HKVEREEG ELSPTESCEQ ENFEVYKENG LEPVQKLPDN EISNTDREPK EGACGTEAVT RSNALPEDDD NKITQKLSEG DE-NASKFIV SASKFG
Q9XIK6-PAH4-At.   ----HELEEG ELSPTASREQ SNFEVNGQNA FKPLQKVTDN VRSNKDKQSC DKKGAKN--- -----KTRAE DDKQENCHKL SE-NNKTASE MLVSGT
O48686-PAH4-At.   --SKVEREEG ELSPNGDFEE DNFAVYAKTD FETFSKANDS TGNNISGDRS REGEPSCLET RAENDAEGDE NAARSSEDSR NE-YENGDVS GTESGG
O04539-PAH4-At.   --AKVEREEG ELSPTGDFEE DNYAVHGEND MEALSKSKEN DATADDASAP ---------- ----RSSDGS GNTSHNGDVS GT-DSGDGED CYREDD
Q658A2-PAH4-Os.   --PKVEREEG ELSPNGDFEE DNFAPFEDGA VDGVSKAKEG STSRPFQGRS GEAQPS-CAE AAGENDADAD DEGEESAQRS TE-DSENASE GGEDAS
Q9SRH9-PAH4-At.   --SRNEKEEG ELSPNGD-FE DNFGVYKDHG VKSTSKPENS AEAEVEAD-- ---------- ----AEVENE DDADDVDSEN AS-EASGTES GGDVCS
Q9LFQ3-PAH4-At.   --SKIEKEEG ELSPVGDSE- DNFVVYEDRE LKATAKTEHS VEAEGEN--- ---------- -----DEDAD DEDGDDASEA GE-DASGTES IGDECS
O01319-PAH2-Ce.   SSELAAHTLN IIELLKKSFL ARPTKLVDFM TFIDFFMSDQ Q--------- ---------- ---YKKDMEK LRKDDEDDEI EE-NEKIEVD DVPGPS
Q61CX5-PAH2-Cb.   PEDVSAQTLE IVRLVTEAFS ARPDRTTSFL TFMDMYLSTT KFYIQ----- ---------- ----KKKGQK MDGGDARDEE MP-GPAPPPQ QVRPPV
Q8H6E1-PAH1-Mp.   -TRSCRVREY LDPELKNLKV KWNFARHGTC PPDARSCAGA RG-------- ---------- ---GKTLSLL NRLASKILFQ FP-AVFIEVS CVSEGR
```

Figure 5.9A Multiple sequence alignment generated by ClustalW of all PAH domains from 35 proteins. Homologues were identified using the Pfam database of protein families (Bateman et al., 2004). Redundant sequences were removed. The alignment was made in three steps: *i*) excluding the most divergent sequences, a multiple alignment was made for every PAH family (PAH1, PAH2, etc.); *ii*) the four family alignments were added together stepwise by a profile alignment; *iii*) divergent sequences were added to the resulting alignment by a sequence alignment which was further optimized manually. Sequences are labeled by their SwissProt accession number, an identifier for the PAH domain and the organism abbreviation. Top row displays the sequence used in the present study in which the helical residues are shown in bold type, the structural role of residue (c=buried residues in the core of the protein, i=residues in the interface between two helices) and the involvement in the binding pocket (h = only intermolecular contacts in the HBP1-complex, m = only intermolecular contacts in the Mad1-complex, + = intermolecular contacts in both complexes).

Chapter 6.

Hugo van Ingen
Geerten W. Vuister
Sybren Wijmenga
and Marco Tessari

Journal of the American Chemical Society
(2006), **128**, p. 3856-3857

CEESY: characterizing the conformation of unobservable protein states

Acknowledgments. We would like to thank prof. dr. Cees W. Hilbers for his stimulating interest, and the Netherlands Organization for Scientific Research, NWO, for financial support (grant JC 99-03).

Protein conformations that are only marginally populated often play important roles as intermediate states in many processes such as ligand binding, enzyme catalysis, allostery and protein folding. An NMR method is presented that can give valuable information about the structure of these 'excited states' by measuring the relative position of exchanging excited and ground state resonances using a single 2D spectrum. This new approach can be applied to any nucleus, which will facilitate a complete structural characterization of these states.

Introduction

Many biological processes such as ligand binding, enzyme catalysis, allosterics and protein folding involve conformational rearrangements on a micro- to millisecond time scale to transient protein states with a low population, also termed excited states (Mulder, et al., 2001; Volkman et al., 2001; Eisenmesser et al., 2002; Korzhnev et al., 2004, ; Dyson and Wright, 2005). These unobservable minor states can only be studied indirectly by exploiting their continuous interconversion into the major, or ground, state. Kinetic and thermodynamic parameters of this exchange process can be determined along with the frequency difference between excited and ground state for a number of backbone and side chain nuclei using well-established relaxation dispersion techniques (Palmer et al., 2001; Akke, 2002; Palmer, et al., 2005; Ishima and Torchia, 2003; Ishima et al., 2004; Lundström and Akke, 2005). These frequency differences are of great interest because they allow the structural characterization of the excited state, provided their sign is known (McElheny et al., 2005; Beach et al., 2005; Grey et al., 2003; Eissenmesser et al., 2005). However, only the absolute value is accessible using relaxation dispersion methods. Experiments to measure the sign have so far only been designed for backbone ^{15}N-nuclei (Skrynnikov et al., 2002; Trot and Palmer, 2002; Korzhnev et al., 2003; Korzhnev et al., 2005).

Here, we present a new method, named CEESY[†] (for Chemical Exchange to Excited States spectroscopY), to establish the sign of the frequency difference between ground and excited state signals using a single 2D

[†] pronounce as /'keisi:/ (International Phonetic Alphabet notation).

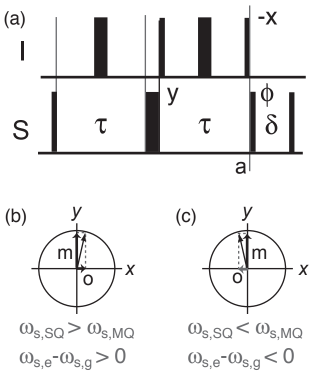


Figure 6.1. (a) Diagram of the CEESY pulse sequence element. Narrow (wide) bars represent 90° (180°) pulses along the *x*-axis unless indicated otherwise. (b,c) In the presence of exchange to an excited state, the echo formed at time point *a* is rotated. The relative sign of the orthogonal (o) and main (m) component matches the sign of the frequency separation between excited and ground state. Passive heteronuclear *J*-couplings involving either spin *I* or *S* are refocused. Passive homonuclear couplings introduce a scaling factor $\cos(\pi J_{S\text{-}S}2\tau)$ that identically affect the orthogonal and main component. Double antiphase terms $4I_zS_xS_z$ and $4I_zS_xSy$, generated by homonuclear coupling $J_{S\text{-}S}$, are removed by the *z*-filter. Cross-correlation effects are suppressed by the 180° pulses on spin *I*.

spectrum. Importantly, this method is not limited to ^{15}N, but can also be applied to most relevant spins in bio-organic materials and will, therefore, allow for a more complete spectroscopic characterization of the excited state.

Pulse sequence design

The essential pulse sequence element of the CEESY experiment involves two heteronuclear spins *I* and *S* (Figure 6.1a). It consists of a modified spin-echo sequence comprised of two periods of duration τ, separated by a refocusing 180° pulse on spin *S*. Single quantum *S* spin coherence evolves in the first τ period and *IS* multiple quantum coherence in the second τ period. In the absence of exchange, the chemical shift evolution of spin *S* is completely refocused and a perfect echo is obtained at point *a*. Conversely, chemical shift refocusing is only partially achieved in presence of exchange, as a result of a difference in the exchange induced shift for SQ and MQ coherences (Skrynnikov et al., 2002) and a small net rotation of *S* coherence at point a results (Figure 6.1b,c). Thus, a large, main component (projection on echo-axis) and a small, orthogonal component (projection on orthogonal axis) are present at the end of this modified spin echo.

The sign of this rotation is determined by the sign of the Larmor frequency difference between excited and ground state resonances in most cases of practical interest, i.e. fast-to-intermediate two-site exchange (Skrynnikov et al., 2002). Observation of identical signs for the two components can therefore be directly translated into an increased Larmor frequency in the excited state ($\omega_{S,\mathrm{e}} > \omega_{S,\mathrm{g}}$), while opposite signs denote a decreased Larmor frequency in the excited state with respect to the ground state ($\omega_{S,\mathrm{e}} < \omega_{S,\mathrm{g}}$). The signs of the two components can readily be obtained by a proper choice of the phase f of the 90° pulse on spin *S* at time point *a* (Figure 6.1a). The CEESY building block is robust with respect to interferences from *J*-couplings and cross-correlated relaxation. The requirement for two-site exchange is not so restrictive. In the case of multi-site exchange, meaningful average spectral information on the conformations in the ensemble of excited states can be obtained (Palmer et al., 2001).

Results and Discussion

We have applied the approach outlined above to determine the sign of the frequency difference for the amide proton ^{1}H$_\mathrm{N}$ ($\Delta\omega_\mathrm{HN} = \omega_\mathrm{HN,e} - \omega_\mathrm{HN,g}$) and the backbone ^{15}N nuclei ($\Delta\omega_\mathrm{N} = \omega_\mathrm{N,e} - \omega_\mathrm{N,g}$). CEESY experiments for ^{1}H$_\mathrm{N}$ and ^{15}N were tested on the 12 kDa PAH2 protein domain, for which fast-to-intermediate conformational exchange has been ob-

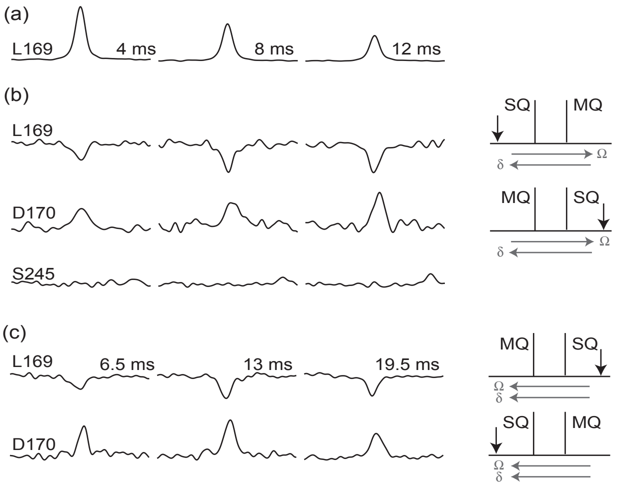


Figure 6.2. Representative results of the $^{1}H_{N}$- (a,b) and ^{15}N-CEESY (c), showing slices through the peak maximum along the acquisition axis of the 2D spectrum. (a) main component of L169 at three echo delay values (indicated above the traces). (b) orthogonal component of L169 (decreased $\omega_{HN,e}$), D170 (increased $\omega_{HN,e}$) and S245 (no exchange). (c) orthogonal components of L22 (decreased $\omega_{N,e}$) and D23 (increased $\omega_{N,e}$). Stick spectra on the right indicate the corresponding relative positions of the ground state SQ- and MQ-peaks and the excited state resonance (arrow). Directions of the frequency and chemical shift axis are shown with grey arrows. Data acquired on 0.9 mM uniformly $^{13}C/^{15}N$ labeled PAH2 sample at 11.7 T, 298K and a room temperature probe head.

served between a structured form and a low-populated, partially unfolded state (Chapter 5). The CEESY building block was inserted in a regular sensitivity enhanced HSQC and for both the main and the orthogonal component a 2D ^{15}N, ^{1}H correlation spectrum was recorded.

Positive and negative peaks were found in the 2D spectrum of the orthogonal component for both $^{1}H_{N}$ and ^{15}N nuclear spins. Representative results are shown in Figure 6.2. Starting with the $^{1}H_{N}$-CEESY, panel (a) shows the main component of the echo for residue L169 in three spectra recorded with increasing duration of the spin echo delay τ. These 1D traces demonstrate the decay of signal intensity due to relaxation and exchange. In panel (b) the traces for orthogonal components of residues L169, D170 and S245 are plotted. For L169 a negative signal is observed, whose magnitude increases with increasing spin echo delay. Likewise, an increasing positive signal is seen for D170. For S245, no signal is observed, implying negligible chemical exchange for the $^{1}H_{N}$ nucleus of this residue. Panel (c) displays two typical results of the ^{15}N-CEESY. A clear, negative signal is observed for L169, and a positive peak for D170.

As discussed above, opposite signs for the two components imply a smaller Larmor frequency for the excited state with respect to the ground state, while identical signs indicate a larger Larmor frequency of the excited state. To translate this into a chemical shift difference, one has to consider that the direction of the frequency axis runs in opposite directions for ^{1}H and ^{15}N as it depends on the sign of the gyromagnetic ratio (Levitt, 1997). Thus, the $^{1}H_{N}$ and ^{15}N resonances of L22 (δ_{HN} 7.3 ppm; δ_{N} 126.0 ppm) are shifted to higher and, respectively, lower ppm values in the excited state. The opposite behavior was found for D23 (δ_{HN} 8.7 ppm; δ_{N} 113.1 ppm), with $^{1}H_{N}$ and ^{15}N resonances shifted to respectively lower and higher ppm values

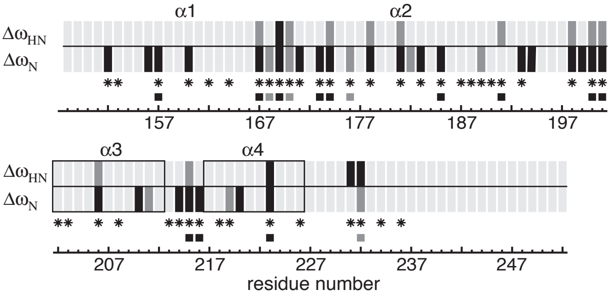


Figure 6.3. Resulting signs (black: negative; dark gray: positive) of $\Delta\omega_{HN}$ (top row) and $\Delta\omega_N$ (bottom row) between ground and excited state. Boxes indicate helices. Stars indicate residues with significant dispersion of ^{15}N-R$_2$ rates (Chapter 5). Signs of $\Delta\omega_N$ from the H(S/M)QC method are shown with squares.

in the excited state. For both residues these results indicate a resonance shift towards the random coil values, consistent with a less structured excited state (see Chapter 5).

The sign of $\Delta\omega_{HN}$ and $\Delta\omega_N$ was classified using the 2D spectrum of the orthogonal component recorded with the longest spin-echo delay (i.e. τ is 12 ms for ^{1}H$_N$ and 19.5 ms for ^{15}N). Taking a conservative approach, only those signals were considered for which the spin-echo was rotated by more than 1^o and for which the intensity, $|I_{\mathrm{ortho}}|$, exceeded three times the noise-level. The maximum rotation observed, defined as arctan($I_{\mathrm{ortho}}/I_{\mathrm{main}}$), was -5 ± 1^o for the ^{1}H$_N$-CEESY and +19 ± 2^o for the ^{15}N-CEESY. Out of 72 non-overlapping resonances, 15 $\Delta\omega_{HN}$ and 34 $\Delta\omega_N$ signs could be classified. The latter accounts for nearly all residues with significant dispersion of their ^{15}N-transverse relaxation rates (Figure 6.3). Notably, even for the fast-relaxing amide proton a significant number of $\Delta\omega$ signs were determined. Interestingly, $\Delta\omega_N$ and $\Delta\omega_{HN}$ have opposite signs for 10 out of 13 possible comparisons.

To validate the CEESY method, the sign of $\Delta\omega_N$ was also determined using the H(S/M)QC approach, as proposed by Skrynnikov et al. (2002). A triplicate set of interleaved HSQC and HMQC spectra was recorded and using a threshold of 0.3 Hz (Skrynnikov et al., 2002), 16 signs could be classified, which were found in perfect agreement with the outcome of the ^{15}N CEESY experiment (Figure 6.3). Only 9 out of these 16 residues showed differences in ^{15}N peak position beyond 1 Hz, with a maximum of 3.2 Hz and an average of 1.1 Hz (taking absolute values). These small differences in peak position are related to the low population (ca. 1%) of the excited state (see Chapter 5). Since the sensitivity of the CEESY experiment is primarily limited by signal-to-noise and spectral analysis is less hampered by signal overlap, a more complete classification of the sign of the shift difference $\Delta\omega$ was possible, even in this case of extremely skewed populations.

Conclusion

In conclusion, we propose a new method to determine the position of unobservable excited state resonances relative to the observed resonances using a single 2D spectrum. The obtained information is crucial for the structural characterization of excited states. Our CEESY experiment can be applied to proteins, nucleic acids or any system in fast-to-intermediate two-site exchange. Here, we have reported the experimental determination of the sign of $\Delta\omega$ of the backbone nitrogen and, for the first time, of the amide proton. Importantly, this approach can be extended to different nuclear species. We anticipate that application to other nuclei, such as the C$_\alpha$ and C$_\beta$ of proteins, will provide spectroscopic information to characterize the structure of unobservable protein conformations in more detail. Ultimately, combination of relaxation-dispersion and CEESY experiments applied to multiple backbone nuclei will allow main chain structure determination of these excited states.

Materials and Methods

Diagrams of the pulse sequences used to measure the sign of $\Delta\omega_{HN}$ and $\Delta\omega_{N}$ are shown in Figure 6.4 (Appendix). CEESY delay τ was set to 4, 8 and 12 ms for the ^{1}H experiment and 6.5, 13 and 19.5 ms for the ^{15}N experiment. In the ^{1}H-CEESY experiment, 192 transients per FID were recorded for the main component and 768 for the orthogonal component (total acquisition times roughly 12/49 hours for main/orthogonal component per τ value). The spectral width of the indirect dimension was set 1200 Hz and 92 complex points were recorded. In the ^{15}N-CEESY experiment, the number of transients per FID was 80-80-320 for the main component and 640-640-1280 for the orthogonal component for τ values 6.5-13-19.5 ms (total acquisition times 4.5/41-4.5/41-18/82 hours for main/orthogonal component). The spectral width of the indirect dimension was set 1200 Hz and 100 complex points were recorded. Final data size after zerofilling was in both experiments 4Kx4K points. Data was recorded at 500 MHz proton frequency and 298 K with a room temperature probe-head.

To ensure accurate results, a ‘zero-test’ was run by setting τ to a small value such that no significant echo rotation will occur (1 ms / 384 transients per FID for the ^{15}N-experiment and 1.6 ms / 512 transients per FID for the ^{1}H-experiment). Proper setting of the phase of the selection pulse (ϕ_6 in the ^{1}H- and ϕ_4 in the ^{15}N-experiment) was verified experimentally as follows (Levitt, 1997; Levitt and Johannessen, 2000; Roehrl et al., 2005). The CEESY element in the ^{1}H experiment is composed of MQ-evolution followed by SQ-evolution. An unbalance was created in the echo by increasing the MQ-period with 230 ms. The echo rotation is then dominated by the timing unbalance (further ensured by using a short τ-value): $|\Omega\tau_{MQ}| > |\Omega\tau_{SQ}|$. Phase ϕ_6 was set such that the sign of the orthogonal and main component are opposite for signals with a positive offset frequency (i.e. when the main component is phased positive, the sign of the orthogonal component corresponds with $\Omega\tau_{SQ} - \Omega\tau_{MQ} < 0$). In the ^{15}N-experiment, the CEESY element is composed of SQ-evolution followed by MQ-evolution. An unbalance was created in the echo by increasing the SQ-period with 200 ms. The echo rotation is again dominated by the timing unbalance: $|\Omega\tau_{SQ}| > |\Omega\tau_{MQ}|$. Phase ϕ_4 was set such that the sign of the orthogonal and main component are identical for signals with a positive offset frequency (i.e. when the main component is phased positive, the sign of the orthogonal component corresponds with $\Omega\tau_{SQ} - \Omega\tau_{MQ} > 0$).

Appendix

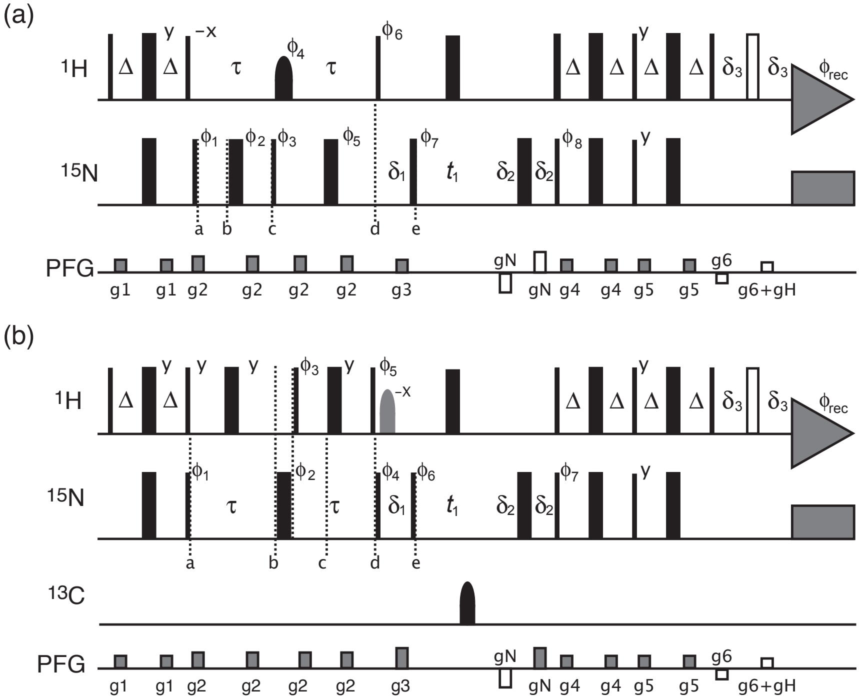


Figure 6.4. Pulse sequence diagrams of the ^1H_N-CEESY experiment to determine sign of $\Delta\omega_{HN}$ (a) and $\Delta\omega_N$ (b). Unless indicated otherwise, all narrow and wide closed bars represent 90° and 180° pulses along the *x*-axis, respectively. Rectangular ^1H (^{15}N) pulses have RF-field strengths (γB_1) of 32.4 kHz (6.1 kHz). All ^{15}N-pulses are centered at 120.0 ppm. In the ^1H-CEESY, all ^1H pulses during the INEPT and the CEESY block are centered on the amide region (8.53 ppm), during the *z*-filter the carrier offset is moved to 4.82 ppm. In the ^{15}N-CEESY, all ^1H pulses are centered on 4.82 ppm. The shaped pulse ϕ_4 in the ^1H-CEESY is a REBURP-shaped refocusing pulse of 2.03 ms with a peak RF-field strength of 3.0 kHz. The proton shaped pulse after ϕ_5 in the ^{15}N-CEESY is a sinc-shaped water-flipback pulse of 1.76 ms. The open bar represents the 3-9-19 implementation of the WATERGATE water suppression technique. The 180° carbon pulse represents a sech-shaped pulse centered at 116 ppm of duration 500 μs and a peak RF-field strength of 4.1 kHz. ^{15}N WALTZ-16 decoupling during acquisition employs a 1 kHz field. Delay Δ was set to $\frac{1}{4}J_{NH}$ (2.65 ms), δ_1 was set to 5.3 (3.5) ms, δ_2 and δ_3 to 1 ms. Recycle delay was 1.1 sec. Phase cycle ^1H-CEESY (Varian Inova nomenclature): $\phi_1=\{x,-x\}4$; $\phi_2=\{x,y\}8$; $\phi_3=\{x,-x\}16$; $\phi_4=\{x,y\}$; $\phi_5=\{x,-x\}64$; $\phi_6=\{y,-y\}2$; $\phi_7=\{x,-x\}32$; $\phi_8=-x$; $\phi_{rec}=\{x,-x,-x,x, -x,x,x,-x, -x,x,x,-x, x,-x,-x,x, -x,x,x,-x, x,-x,-x,x, x,-x,-x,x, -x,x,x,-x, -x,x,x,-x, x,-x,-x,x, x,-x,-x,x, -x,x,x,-x, x,-x,-x,x, -x,x,x,-x, -x,x,x,-x, x,-x,-x,x\}$. Phase cycle ^{15}N-CEESY (Varian Inova nomenclature): $\phi_1=\{x,-x\}8$; $\phi_2=\{x,y,-y,-x\}2$; $\phi_3=\{y,-y\}16$; $\phi_4=\{x,-x\}$; $\phi_5=\{y,-y\}32$; $\phi_6=x$; $\phi_7=-x$; $\phi_{rec}=\{x,-x,-x,x, -x,x,x,-x, -x,x,x,-x, x,-x,-x,x, -x,x,x,-x, x,-x,-x,x, x,-x,-x,x, -x,x,x,-x, -x,x,x,-x, x,-x,-x,x, x,-x,-x,x, -x,x,x,-x, x,-x,-x,x, -x,x,x,-x, -x,x,x,-x, x,-x,-x,x\}$. All ^1H pulses executed with phase *y* or *-y* should be inverted to obtain the physically correct nutation axis (Roehrl et al., 2005). P- and N-type FIDs are recorded for the ^1H-CEESY (^{15}N-CEESY) by adding 180° to phase $\phi8$ ($\phi7$) and inverting gradient g_N. Axial peaks are shifted to edges of the spectrum by adding 180° to $\phi7$ ($\phi6$) and ϕ_{rec} every other increment. The orthogonal component of the echo formed at the end of the CEESY element is selected by adding 270° to $\phi6$ ($\phi4$). In the ^1H-CEESY, the phase of all proton pulses after this pulse and of the receiver are simultaneously incremented 90° to control the state of the water magnetization. Pulsed field gradients were hyperbolic sine shaped and employed along the *z*-axis. Duration of all gradients was set to 1 ms except gradient g2 that was set to 350 ms. Gradient strengths were (in % of maximum strength, 50 G/cm) for the ^1H-CEESY (^{15}N-CEESY): g1=3 (3); g2=20 (25); g3=7 (30); g4=6 (6); g5=9 (9); g6=-24 (-24); gN=60 (60); gH=12.135 (12.161).

Table 6.1A.
Results of the sign of $\Delta\omega_{HN}$ and $\Delta\omega_{N}$ using the CEESY method and comparison with the H(S/M)QC approach.

resnr.[a]	$^{1}H_{N}$ or ^{15}N data	chemical shift (ppm)	sign[b]	SNR orthogonal component[c]	rotation echo (deg)[d]	error rotation (deg)[d]	SQ-MQ (Hz) H(S/M)QC[e]	error SQ-MQ (Hz) H(S/M)QC[e]	Δω dispersion analysis (ppm)[f]	random coil - observable (ppm)[g]
5	HN	8.27	< th.	14.8	0.882	0.057				
	N	122.93	-	-13.5	-1.025	-0.074	-0.14	0.04	1.63	-3.63
6	HN	8.51	< th.	11.6	0.670	0.057				
	N	122.13	< th.	-3.5	-0.275	-0.080	0.17	0.03	1.99	
9	HN	8.66	< th.	2.8	0.762	0.269				
	N	120.01	-	-9.4	-2.628	-0.281	-0.30	0.13		-1.01
10	HN	8.36	< th.	3.7	0.796	0.212				
	N	124.80	-	-8.8	-2.325	-0.269	-0.38	0.06	1.95	0.20
13	HN	8.10	< th.	-0.7	-0.155	-0.223				
	N	123.90	-	-4.4	-1.375	-0.309			1.98	-3.00
15	HN	8.21	< th.	0.8	0.138	0.166				
	N	117.44	< th.	-1.5	-0.338	-0.223	0.01	0.08	1.77	
17	HN	7.64	< th.	1.2	0.413	0.355				
	N	120.14	< th.	-2.8	-1.375	-0.493	-0.44	0.18	1.80	
20	HN	8.11	+	6.9	1.747	0.252				0.29
	N	123.45	-	-6.9	-2.376	-0.349	-1.25	0.17	2.63	-2.24
21	HN	7.36	< th.	-2.3	-1.071	-0.475				
	N	113.60	+	7.4	6.005	0.822	1.18	0.22	3.69	7.11
22	HN	7.25	-	-14.2	-4.141	-0.292				1.03
	N	126.05	-	-11.5	-6.153	-0.542	-1.04	0.02	5.87	-3.65
23	HN	8.68	+	10.9	2.628	0.240				-0.12
	N	113.13	+	9.2	3.234	0.355	0.89	0.22	3.15	5.97
24	HN	7.77	< th.	4.6	0.550	0.120				
	N	117.18	-	-9.2	-1.610	-0.178	-0.23	0.08	1.04	0.92
26	HN	10.31	< th.	2.9	1.409	0.487				
	N	121.59	-	-11.6	-8.093	-0.717	-1.36	0.27	2.73	-1.39
27	HN	8.17	+	4.1	1.426	0.349				0.00
	N	123.23	-	-8.6	-4.352	-0.509	-0.86	0.21	2.61	-2.83
29	HN	8.42	< th.	4.6	0.842	0.183				
	N	117.32	+	5.0	1.255	0.252	0.36	0.03	1.29	3.88
31	HN	8.26	+	8.7	1.879	0.218				0.05
	N	123.39	-	-6.5	-2.022	-0.309	-0.03	0.06	1.77	-2.69
34	HN	7.99	+	6.0	1.358	0.223				0.18
	N	122.26	-	-5.4	-1.713	-0.321	-0.01	0.22	1.48	-1.86
35	HN	7.81	< th.	0.7	0.281	0.418				
	N	119.34	+	3.9	2.159	0.561				3.06
36	HN	8.64	< th.	1.1	0.831	0.756				
	N	117.65	-	-3.2	-2.142	-0.676	-0.21	0.20	1.99	0.45
38	HN	8.28	< th.	-1.6	-0.487	-0.315				
	N	122.17	-	-5.2	-2.148	-0.412	-1.21	0.17	1.27	-1.27
40	HN	7.98	< th.	2.3	0.355	0.155				
	N	119.49	< th.	-2.1	-0.453	-0.218	-0.05	0.17	1.44	
41	HN	8.06	< th.	5.0	0.814	0.166				
	N	118.93	< th.	-3.8	-0.888	-0.235	-0.02	0.11	1.50	
42	HN	7.88	< th.	3.6	0.791	0.218				
	N	117.72	+	5.1	1.570	0.304	0.33	0.33	2.10	2.78

Table 6.1A.
Results of the sign of $\Delta\omega_{HN}$ and $\Delta\omega_N$ using the CEESY method and comparison with the H(S/M)QC approach.

resnr.[a]	^1H_N or ^{15}N data	chemical shift (ppm)	sign[b]	SNR orthogonal component[c]	rotation echo (deg)[d]	error rotation (deg)[d]	SQ-MQ (Hz) H(S/M)QC[e]	error SQ-MQ (Hz) H(S/M)QC[e]	$\Delta\omega$ dispersion analysis (ppm)[f]	random coil - observable (ppm)[g]
43	HN	7.73	< th.	2.1	0.401	0.195				
	N	120.06	< th.	-0.5	-0.132	-0.252	0.12	0.11	1.86	
44	HN	8.23	< th.	3.2	1.919	0.596				
	N	119.17	< th.	-1.8	-1.237	-0.693	-1.21	0.21	2.32	
46	HN	8.42	< th.	3.3	0.796	0.241				
	N	123.92	-	-7.1	-1.844	-0.263	-0.25	0.05	1.95	-2.32
47	HN	8.65	< th.	1.7	1.873	1.083				
	N	110.58	-	-5.0	-6.045	-1.233	-0.85	0.69		-3.08
51	HN	8.33	< th.	3.8	1.381	0.361				
	N	124.17	-	-8.2	-4.021	-0.492	-0.56	0.34	3.69	-2.97
52	HN	7.81	< th.	0.6	0.229	0.390				
	N	109.84	-	-4.4	-1.919	-0.430	0.19	0.31		-2.34
53	HN	8.53	< th.	3.2	1.392	0.435				
	N	121.10	-	-12.5	-6.124	-0.497	-0.64	0.17	1.80	-0.80
54	HN	8.95	+	10.3	3.028	0.298				-0.52
	N	120.48	-	-7.5	-3.622	-0.486	-1.06	0.32	4.12	-4.98
55	HN	9.03	< th.	5.6	0.940	0.172				
	N	121.33	< th.	2.9	0.607	0.212	0.26	0.01	1.03	
56	HN	8.70	< th.	5.7	0.917	0.160				
	N	118.48	< th.	-4.0	-0.882	-0.223	-0.07	0.05	1.33	
59	HN	8.27	+	7.6	1.243	0.166				0.04
	N	119.42	-	-6.5	-1.472	-0.229	0.06	0.14	1.74	1.28
61	HN	7.99	< th.	4.2	0.653	0.155				
	N	120.56	< th.	2.4	0.516	0.218	0.16	0.08	1.44	
63	HN	8.47	< th.	2.5	0.321	0.126				
	N	122.17	-	-6.6	-1.163	-0.178	-0.22	0.00		1.28
64	HN	7.46	< th.	1.8	0.630	0.349				
	N	114.33	+	6.2	2.862	0.464	0.58	0.38		4.67
66	HN	8.15	< th.	2.3	0.877	0.372				
	N	114.85	< th.	-1.3	-0.693	-0.539	-0.32	0.27	1.59	
67	HN	7.02	< th.	-0.4	-0.103	-0.252				
	N	120.83	-	-3.8	-1.180	-0.315	0.02	0.30	1.15	0.37
68	HN	9.35	+	4.2	4.853	1.172				-0.94
	N	115.26	-	-6.7	-10.869	-1.693	-2.13	1.21	4.46	-7.76
69	HN	8.50	< th.	2.9	0.401	0.138				
	N	120.50	-	-14.4	-2.960	-0.206	-0.50	0.01	1.04	0.00
71	HN	8.98	< th.	7.8	0.842	0.109				
	N	119.29	< th.	-6.5	-0.928	-0.143	-0.05	0.05	1.18	
72	HN	7.78	< th.	-1.1	-0.183	-0.172				
	N	119.60	+	5.4	1.318	0.246	0.26	0.05	2.66	2.80
73	HN	8.05	< th.	5.5	0.814	0.149				
	N	119.98	-	-7.2	-1.518	-0.212				2.42
76	HN	7.85	-	-5.8	-2.554	-0.441				0.46
	N	121.44	-	-4.9	-4.067	-0.841	-0.80	0.19	6.68	-0.74
79	HN	7.51	< th.	-0.6	-0.854	-1.490				
	N	120.40	< th.	-2.0	-6.113	-3.077	-0.51	0.62	5.51	

Table 6.1A.
Results of the sign of $\Delta\omega_{HN}$ and $\Delta\omega_{N}$ using the CEESY method and comparison with the H(S/M)QC approach.

resnr.[a]	^1H_N or ^{15}N data	chemical shift (ppm)	sign[b]	SNR orthogonal component[c]	rotation echo (deg)[d]	error rotation (deg)[d]	SQ-MQ (Hz) H(S/M)QC[e]	error SQ-MQ (Hz) H(S/M)QC[e]	$\Delta\omega$ dispersion analysis (ppm)[f]	random coil - observable (ppm)[g]
84	HN	7.97	-	-5.6	-1.226	-0.223				0.39
	N	116.05	< th.	2.2	0.665	0.309	0.00	0.11	3.95	
85	HN	7.58	-	-4.5	-5.148	-1.171				0.81
	N	119.70	+	10.8	19.565	2.119	3.22	0.90	6.23	1.50
87	HN	8.04	< th.	-1.3	-0.395	-0.315				
	N	121.98	< th.	-3.0	-1.232	-0.412	0.16	0.23	2.32	
89	HN	7.97	< th.	-2.2	-0.842	-0.384				
	N	112.74	< th.	-2.3	-0.607	-0.269	0.17	0.09	2.69	

[a] only those residues are shown for which either the sign of $\Delta\omega_{HN}$ or $\Delta\omega_{N}$ was determined or significant ^{15}N-R_2 dispersion was observed.

[b] sign based on the CEESY-experiment. Criteria: signal-to-noise > 3 and echo rotation > 1^o. Residues for which no sign could be determined are indicated by “< th.”

[c] Signal-to-Noise Ratio for the orthogonal component observed in the spectra with the longest echo delay (12 ms for ^1H_N and 19.5 ms for ^{15}N).

[d] echo rotation (α) for τ = 12 ms (^1H_N) or 19.5 ms (^{15}N) was calculated using the relation: $\alpha = \arctan(I_{ortho}/I_{main})$, in which I_{ortho} and I_{main} are the intensities in the orthogonal and main component, both rescaled to 1 scan to compensate for the difference in acquisition time. Errors are based on the noise-level.

[e] difference in peak position along the ^{15}N-dimension between HSQC and HQMC spectra. Average value and standard deviation of three measurements. HSQC and HMQC spectra were recorded interleaved at 500 MHz proton frequency; spectral width of the indirect dimension was set to 1200 Hz; 92 complex points; 64 transient per FID.

[f] magnitude of the ^{15}N chemical shift difference between excited and ground state as determined by a global fit of CPMG ^{15}N R_2-relaxation dispersion data. Dispersion curves were recorded at 298K and 500, 600 and 800 MHz proton frequency. Curves were fit to the Carver-Richards-Jen (Carver and Richards, 1972; Jen, 1978) all time scale approximation describing the apparent R_2 as function of the pulse spacing in CPMG experiments.

[g] chemical shift difference between the random coil shift and the ground state ^1H_N or ^{15}N shift. Random coil shifts based on the Wright & Dyson values deposited in the BMRB database (Schwarzinger et al., 2000).

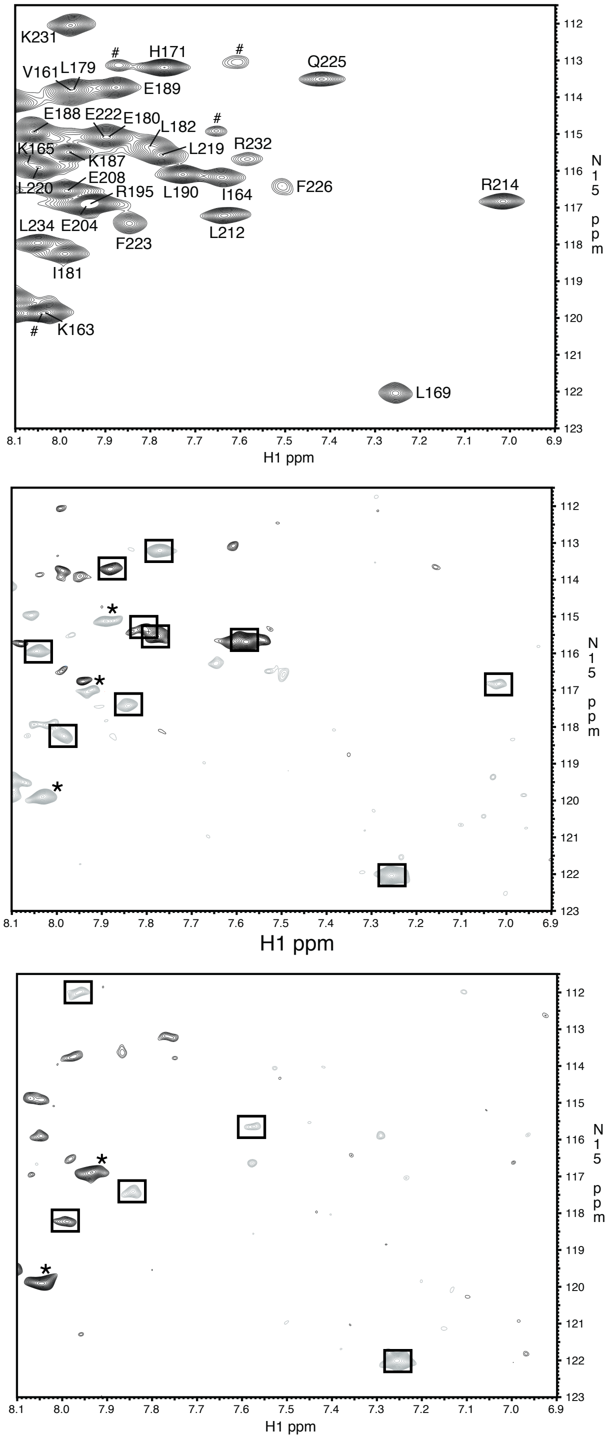


Figure 6.5A. Sections of the 2D spectra of a reference ^{1}H,^{15}N-HSQC spectrum (top panel), the orthogonal component spectrum of the ^{15}N-CEESY experiment (CEESY delay τ is 19.5 ms; middle panel) and the orthogonal component spectrum of the ^{1}H-CEESY experiment (CEESY delay τ is 12 ms; bottom panel). Positive and negative peaks are colored blue and red, respectively. The orthogonal component spectra are contoured at the noise level. The assignment of the resonances is shown in the top panel. Peaks labeled with a hash (#) result from fragments due to protein degradation. Peaks in the orthogonal component spectra that fulfill the criteria described in the main text (SNR > 3 and $\arctan(I_{\mathrm{ortho}}/I_{\mathrm{main}}) > 1°$), are enclosed in boxes and were used to classify the sign of the frequency difference between the ground and excited state (see also Table S1). Peaks labeled with a star (*) are not considered due to overlap.

Chapter 7.

General discussion

The research described in this thesis explores the relation between the structure, dynamics and function of the PAH2 domain of the co-repressor protein Sin3. It was found that the free PAH2 domain exists predominantly in a folded form where the binding pocket is exposed to the solvent. This folded form is in exchange with a minor conformation which is most likely a partially unfolded state (Chapter 5). Upon binding of its target peptide, derived from the Sin3 Interacting Domain of the repressor protein Mad1, the global fold of the PAH2 domain and the side chain conformation in the binding pocket remains virtually unchanged (Chapters 4 and 5). The interaction between PAH2 and the SID is mainly hydrophobic but also involves a crucial hydrogen bond (Chapter 4). Complex formation quenches the exchange between the folded unbound PAH2 domain and the minor conformation and also reduces the flexibility of the N-terminal part of helix α1 (Chapters 4 and 5). Below, I will discuss the structure of the PAH domain in general and the specificity of the PAH-SID interaction. I propose a model to explain how the homologous PAH1, PAH2 and PAH3 domains can discriminate between their specific target proteins and how multiple repressors can bind to a single member of the PAH domain family. In addition, I suggest a mechanism for complex formation in which conformational selection and the induced fit mechanism are mixed, but are involved in different stages of complex formation. Finally, a number of suggestions for further research is made.

Structure of the PAH domains

In Chapter 5 it is shown that the structure of the PAH2 domain of mSin3B, PAH2^{B}, is highly similar in both the unbound and bound states. The unbound PAH2 domain exists in an so-called 'open' conformation in which the binding pocket is accessible to the solvent. This folded form is in fast-to-intermediate exchange with a minor, partially unfolded conformation. In the bound state, this exchange is no longer observed.

The PAH2 domain of the isoform mSin3A, PAH2^{A}, studied in the laboratories of Eisenman and Radakrishnan, has an overall sequence identity of 61% and a 75% sequence identity for the structured regions. The structure of this PAH2 domain was solved in its bound state complexed to the SID of the repressor Mad1 (residues 6-21; Brubaker et al., 2000) and the SID of HBP1 (residues 358-380; Swanson et al., 2004), but not in its unbound state. The overall fold of PAH2^{A} and PAH2^{B} is the same in all complexes and the side chain configuration of amino acids in their binding pockets is highly similar. One striking difference, however, is the extension of helix α2 in PAH2^{A} by almost two turns compared to PAH2^{B}, which most likely results from the higher intrinsic helix propensity of PAH2^{A} for this region (Chapter 4). In the unbound state, PAH2^{A} is in slow exchange between two, roughly equally populated, conformations as opposed to the fast interconversion to a minor, partially unfolded state for PAH2^B. The limited chemical shift dispersion and relatively low heteronuclear NOE values for PAH2^A indicate the presence of a (partially) unfolded domain (Brubaker et al., 2000). I suggest that PAH2^A is also in exchange between a folded and partially unfolded form. At the moment, it is still unclear how the intriguing difference in the exchange kinetics and thermodynamics in PAH2^A *vs.* PAH2^B is related to the differences in structure and sequence (Chapter 5).

Sequence alignment of all PAH domains of 35 Sin3 proteins from various organisms shows that the PAH1, PAH2 and PAH3 domain have very similar sequences, whereas the sequence of the PAH4 domain is more divergent. Therefore, on the basis of this alignment and the structure of the PAH2 domain, I predicted that the PAH1 and PAH3 domain will have a similar fold as the PAH2 domain, whereas the PAH4 domain will likely have a different fold (Chapter 5). So far, no structural data has been published for the PAH3 or PAH4 domain to confirm this hypothesis. However, the structure of the PAH1 domain of mSin3B complexed to the Sin3 interacting domain of the repressor NRSF/REST (neural restrictive silencing factor/repressor element 1 silencing transcription factor) was published recently and shows the PAH1 domain to be very similar to the PAH2 domain (Nomura et al., 2005). Similar to the PAH2 domain, the PAH1 domain is also folded as a four-helix bundle where the SID is wedged between helices α1 and α2 (see also Figure 7.1). Helices α1, α2 and α4 are, however, a bit shorter in the PAH2 domain. Furthermore, a conserved stretch of residues following helix α4 folds back onto the PAH1 domain. The NRSF/REST-SID contacts residues in the PAH1 domain that nearly all correspond to structurally homologous residues in the Mad1-PAH2 complex.

Specificity of protein-protein interactions

The structures of the Mad1-PAH2, HBP1-PAH2 and NRSF/REST-PAH1 interaction provide an opportunity to gain insight in the molecular determinants of specificity of the PAH-SID interaction, both from the point of view of the PAH domain and from the point of view of the repressor. Although specificity is a familiar term in biochemistry, its definition and its structural requirements are far from trivial.

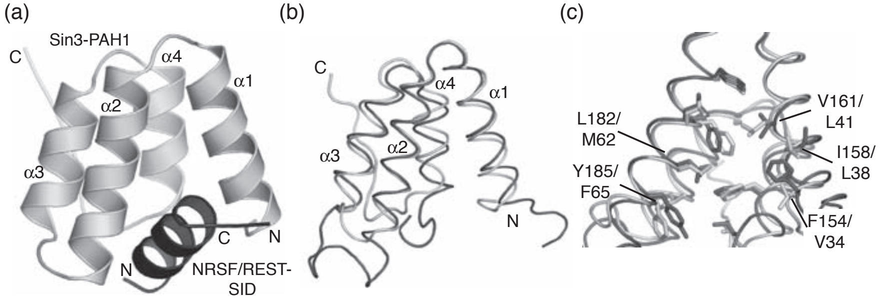


Figure 7.1. Solution structure of the complex between the NRSF/REST-SID and the PAH1 domain of Sin3 (Nomura et al., 2005; PDB code 2CZY). (a) Cartoon representation showing that the two interaction domains are folded as a wedged helical bundle. The SID is shown in dark grey, the PAH1 domain in light grey. (b) structural superposition of the PAH1 (light grey) and the unbound PAH2 domain (dark grey). (c) close-up of (b) showing selected side chains in the binding pocket. Substitutions in the binding pocket are labeled, listing first the PAH2 residue, then the PAH1 residue.

From a thermodynamic perspective, the specificity of an interaction between a protein and a ligand can be inferred by comparing the change in free energy (ΔG) of this interaction with the ΔG of interactions with other ligands, i.e. by calculating the $\Delta(\Delta G)$. Specificity can also be related to kinetics as in the study of enzyme kinetics. Here, the specificity of an enzyme is described by its specificity constant for a particular substrate, defined by the ratio of the turnover number kcat and the Michaelis-Menten constant K_M, and describes the reaction rate for the turnover of substrate into product. The specificity of an enzyme can then be assessed by comparing the specificity constants for different substrates.

In the case of protein-protein interactions, a specificity constant cannot be similarly defined. Rather, two proteins are said to interact specifically if *i*) their interaction is not the result of a random association and *ii*) their interaction is selective or competitive, i.e. their ΔG compares favorably with the ΔG of other interactions involving the two proteins. I will address these two points on a structural level by considering the interaction between a protein *A* and a protein *B*. Both proteins can have multiple interaction surfaces, represented as I_A^1, I_A^2, I_A^3 ... I_A^n and I_B^1, I_B^2, I_B^3 ... I_B^m. Each interaction surface is defined by a set of residues that together form a certain sequence motif.

Considering for the moment only the requirement for a non-random association, protein *A* is said to interact specifically with protein *B* and, likewise, protein *B* interacts specifically with protein *A*, if they bind each other exclusively via interaction surfaces I_A^i and I_B^j, where *i* and *j* refer to a particular interaction surface on protein *A* and *B*, respectively (see Figure 7.2a). Here, specificity is the result of the formation of well-defined intermolecular interactions. Hence, at the molecular level, a specific interaction is the result of complementarity between the interaction surfaces of the two proteins, both with regard to the shape of the interface and its chemical character. Assessment of the specificity of an interaction requires knowledge of these intermolecular interactions either by deletion, mutation or structural studies.

The requirement for a competitive interaction arises when we consider the interaction between protein *A* and *B* in the context of a pool of proteins. Let us consider a pool of three proteins *A*, *B*, and *C*. Suppose specific complexes, according to the first requirement, can be formed between protein *A* and

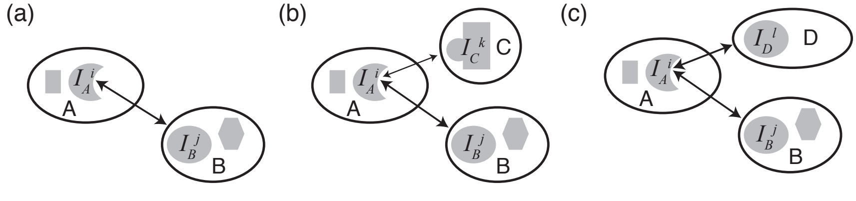


Figure 7.2. Schematic representation of protein-protein interactions. Proteins are represented as ellipsoids. Interaction surfaces are shown as differently shaped grey boxes. (a) a specific interaction between protein *A* and *B* requires the formation of intermolecular interactions between a single, defined interaction surface on protein *A*, I_A^i, and a single defined interaction surface on protein *B*, I_B^j. (b) In this scheme, protein *A* specifically interacts with protein *B* and *C* as it exclusively uses its surface I_A^i to bind to surfaces I_B^j and I_C^k. However, the *AB* complex is preferentially formed as a higher degree of shape complementarity in the *AB* complex compared to the *AC* complex results in a higher binding affinity, indicated by the thicker arrows. (c) As in (b), but now interaction surface I_D^l on protein *D* is similar to surface I_B^j on protein *B* as they are both defined by a conserved sequence motif.

B and protein *A* and *C* (see Figure 7.2b). To predict which complex will dominate in this pool, we need to know the binding affinities (ΔG) of these complexes and the concentrations of all the proteins. Assuming equal concentrations for all proteins, complex *AB* will be predominantly formed if the affinity of *A* for *B* is higher than the affinity of *A* for *C*. Although protein *A* and *C* can form a specific complex, their interaction is too weak to be able to compete with the interaction between *A* and *B*. Thus, it is said that only protein *A* and *B* interact specifically in this case. The requirement for a competitive interaction is connected at a structural level to the first requirement for a non-random association. A higher degree of complementarity between the interaction surfaces (as suggested in Figure 7.2b) will result in a higher binding affinity for the complex, thus resulting in a more specific interaction.

One can distinguish proteins, or more precisely, protein-protein interaction surfaces with low or high specificity; a surface with low specificity will form well-defined complexes with many other proteins, all with comparable affinities, while a surface with high specificity will form a high affinity complex with only one protein, and possibly low affinity complexes with other proteins. Notably, an interaction surface can have a high specificity and still bind to many different proteins if these proteins have similar interaction surfaces defined by a conserved sequence motif (see Figure 7.2c). Hence, specificity is the result of the formation of well-defined, high affinity intermolecular interactions between surfaces defined by unique sequence motifs.

Finally *in vivo*, protein-protein interactions not only depend on the binding affinity and specificity as mentioned above, but are also determined by the localization of the interacting proteins and their relative concentrations.

Specificity of the PAH-SID interaction: the PAH perspective

First, I will discuss the specificity of the PAH-SID interaction in view of the highly similar structures of the PAH1, PAH2 and, presumably, the PAH3 domain. Repressor proteins are reported to bind, *in vivo*, to only to one of the PAH domains. For instance, the repressors Mad1 and HBP1 specifically bind the PAH2 domain (Ayer et al., 1995; Schreiber-Agus et al., 1995; Swanson et al., 2004), whereas the repressor NRSF/REST specifically binds the PAH1 domain (Nomura et al., 2005). However, since the PAH-SID interaction is mainly hydrophobic and the structures of the binding pockets of three PAH domains are similar, one would expect also a substantial interaction with the other PAH domains. In fact, NMR experiments demonstrated that the Mad1 repressor can also bind to the PAH1 domain under NMR conditions (unpublished results).

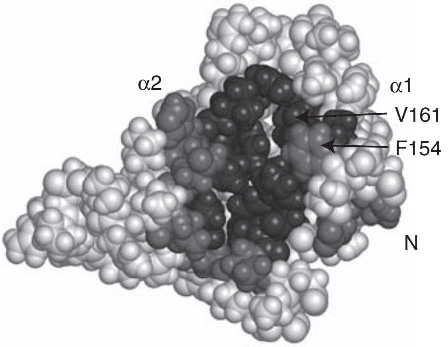


Figure 7.3. The binding pocket between helices α1 and α2 comprises two parts. Side chains of residues in the central part of the binding pocket (shown in black) become completely buried in the complex. They are involved in the general PAH-SID interaction and most of them are conserved between the PAH domains. Side chains of residues at the edge of the binding pocket (shown in dark grey) remain partly accessible to the solvent in the complex, are less conserved and can participate in interactions specifically with one repressor. The N-terminus, helices α1 and α2 and residues F154 and V161 are indicated.

Although this interaction is not the result of a random association, but relies most likely on well-defined intermolecular interactions, it is not a competitive interaction because it has a lower binding affinity compared to the interaction with the PAH2 domain. This shows that subtle differences in the shape and chemical character of the interaction surface between the different PAH domains will significantly affect the binding affinity of the PAH-SID interaction. The weak PAH1-Mad1 interaction is probably not significant *in vivo*, as in that case high affinity PAH1 targets will compete with Mad1 for the PAH1 domain. Furthermore, the PAH2 and PAH1 domain will compete for Mad1.

The specificity of the repressor proteins for one of the PAH domains has to be related to sequence differences between the different PAH domains, resulting in slightly different interaction surfaces. Using the sequence alignment of these domains, residues can be identified that are conserved within one PAH domain but are replaced by a different type of amino acid in the other PAH domains (see Table 7.1). Interestingly, sequence divergence is highest for residues located at the edges of the interaction surface, which remain partly solvent exposed in the complex. In contrast, residues in the central part of the interaction surface are completely buried upon complex formation (see Figure 7.3). Although most of these residues are highly conserved between the different PAH domains, some 'central' residues are not conserved, albeit that the sequence divergence for this group is clearly less than for the 'edge' residues (see listed in Table 7.1).

In an interesting study by le Guezennec et al. (2004), the specificity of the Mad1-repressor for the PAH2 *vs.* the PAH1 domain was investigated by swapping residues of the PAH2 to the PAH1 domain and vice versa. Mutations were identified that resulted in *i*) the loss of PAH2-Mad1 interaction by substituting the corresponding residues of

Table 7.1.
Sequence diversity of binding pocket residues in all PAH domain members.[a]

PAH2[B]-residue	PAH1[b]	PAH2[b]	PAH3[b]
V152	**X**	**V,L**	**X**
E153	**X**	**E,D**	**X**
F154	**V,T[c]**	**F[c]**	**X**
A157	A	A	E,D
I158	L	I	X
V161	L[c]	V[c]	F (L,C)
L182	M	L	L,I
H183	**K**	**Q,H,N**	**N (V,A)**
Q186	**K[c]**	**Q,R[c]**	**S,N (V)**
E202	**T**	**I,E**	**R,K (G,T)**
V205	V	V	L

[a] Residues located at the edges of the interaction surface are shown in bold type, residues in the central part of the interaction surface in normal type. The central part is defined by side chains that become completely inaccessible to the solvent upon complex formation. Side chains of residues at the edge, or rim, of the binding pocket remain solvent accessible.

[b] Most frequent amino acid types are listed that occur at a certain position in the PAH*i* sequences (*i*=1,2,3) (80% cut-off). Residues are listed in order of decreasing occurrence. When some amino acids occur in the majority of sequences, the ones occurring in the minority are placed between round brackets. X denotes any amino acid.

[c] Residues identified as determinants of the specificity of Mad1 for PAH2 *vs.* PAH1 in the study of le Guezennec et al. (2004).

the PAH1 domain and *ii*) the gain of PAH1-Mad1 interaction by substituting residues of the PAH2 domain into the PAH1 domain. Three residues in the PAH2 domain, F154, V161 and Q186, were identified as the critical determinants of specificity, of which residue F154 was found to be most important (see also Table 7.1). Residue F154 is located in the N-terminal half of helix α1 at the edge of the binding pocket (see also Figure 7.3) and is close to Y18 of Mad1. Le Guezennec et al. propose that these residues are involved in a stacking interaction, crucial for the molecular recognition between the PAH2 domain and the Mad1-SID.

The results of le Guezennec et al. (2004) are specific for the case of Mad1-PAH2 *vs.* Mad1-PAH1 interaction: the crucial role of F154 depends on the presence of Y18 in Mad1. If this experiment would be performed for the HBP1 repressor, which lacks an aromatic residue at the corresponding position of Y18, residue F154 would not be identified as it is simply not involved in intermolecular interactions. Instead, the HBP1 repressor uniquely interacts with PAH2 residues V152, E153, and E202 (Swanson et al., 2004), which are also residues unique to PAH2 (see Table 7.1). Thus, these residues might be identified as the determinants of specificity of HBP1 for the PAH2 *vs.* the PAH1 domain.

The identification of residue V161 as a determinant of specificity for the PAH2 domain (le Guezennec et al., 2004) is very interesting as this residue is in the central part of the binding pocket and is involved in intermolecular interaction with not only Mad1, but also HBP1 and presumably all other SIDs. It is the only residue in the central part of the binding pocket that is not structurally conserved between PAH1 and PAH2 (see Figure 7.1c and Table 7.1). Therefore, I propose that the substitution of residue V161 in the PAH2 domain with the more bulky residue L41 at the corresponding position in the PAH1 domain is the only general factor discriminating these two domains. Furthermore, V161 in PAH2 and the corresponding L41 in PAH1 will also be a general factor discriminating the PAH1 and PAH2 domain from the PAH3 domain, where the corresponding residue is mostly a phenylalanine (Table 7.1). In addition, at least two more residues in the central part of the binding pocket of the PAH3 domain will be general factors determining its specificity: the glutamic acid residue at the position corresponding to A157 in PAH2 and the leucine residue corresponding to V205 (see Table 7.1). The presence of a charged residue in the central part of the binding pocket is rather surprising and will certainly be reflected in the SIDs that bind specifically to the PAH3 domain.

In conclusion, residues in the central part are conserved between the different PAH domains and are important for the general PAH-SID interaction. The current data suggest that a few subtle differences in this part between the PAH1, PAH2 and PAH3 domain are general factors determining the specificity of the PAH domain, e.g. L41 of PAH1 *vs.* V161 of PAH2 *vs.* F302 of PAH3. Residues at the edge of binding pocket are not always involved in interaction with all SIDs, but can form unique interactions with a specific repressor. These 'edge' residues can be crucial factors that discriminate between the different PAH domains for specific repressors, like F154 of PAH2 for Mad1. The 'border' residues might not be as important for the PAH3 domain as they appear to be not conserved within this domain (see Table 7.1).

Specificity of the PAH-SID interaction: the SID perspective

Secondly, I will discuss the specificity of the PAH-SID interaction in view of the many different repressor proteins that bind to one of the PAH domains *in vivo*. As the Sin3 protein is a co-repressor, it functions by binding to many repressor proteins. This promiscuity of the PAH domain can either be explained by a low specificity if the in-

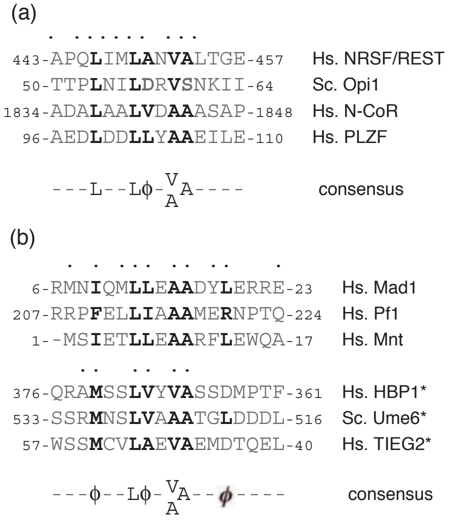


Figure 7.4. Sequence alignment of Sin3 Interacting domains of various repressor proteins. (a). Sequences of SIDs of PAH1 interacting proteins. (b) Sequences of SIDs of PAH2 interacting proteins, aligned with respect to the sequences in (a). Sequences indicated with an asterisk are presented in reverse. Dots above the sequences of NRSF/REST, Mad1 and HBP1 denote residues involved in intermolecular contacts. Consensus sequences are shown below the alignments in (a) and (b); ϕ represents a bulky hydrophobic residue. The SID consensus for the PAH2 interactors contains a bulky hydrophobic residue ϕ (in italic type) that is not strictly conserved. No SID has yet been identified for the PAH3 domain.

teraction surfaces on the SIDs are unrelated or by high specificity of the PAH domain if these surfaces are all defined by a single conserved sequence motif. As is shown in Figure 7.4, the sequence identity between the SIDs of different PAH2 and PAH1 interacting proteins seems to be rather limited and is restricted to the part of the SID that interacts with the central part of the binding pocket of the PAH domain. A conserved pattern of bulky and small hydrophobic residues is observed, consistent with the ‘knob-into-holes’ type of interaction. A recent mutation study stressed the importance of the complementarity of bulky and small residues in the center of the interaction surface as a basis for affinity and specificity of the Mad1-PAH2 interaction (Cowley et al., 2004). The side chains of residues L11, L12, A15 and A16 in the Mad1-SID were shown to complement a pocket formed by residues I158, V161, L179 and L182 in PAH2. Notably, the crucial role of the side chains is further emphasized by the presence of a group of ‘reverse’ binders for the PAH2 domain, where the sequence motif is reversed resulting in a reversal of the helix orientation of the SID, as was first observed for the HBP1 repressor (Swanson et al., 2004).

Although the pattern of bulky and small residues in the consensus sequence confers some specificity, it can be expected that hydrophobic or amphipathic helices with a partial match to the consensus motif, will still bind to the PAH1 or PAH2 domain, albeit with lower affinity. Again, this might not be significant *in vivo*, because of the competition with proteins that bind with high affinity to Sin3. The high binding affinity of proteins such as Mad1 not only results from the shape complementarity as defined by the consensus-motif, but also results from additional intermolecular interactions involving residues outside the consensus-motif (see also Figure 7.4 and Chapter 4).

These additional interactions point to structural adaptivity in the PAH-SID complex. Superposition of the Mad1-PAH2, HBP1-PAH2 and NRSF/REST-PAH1 structures shows that the orientation and position of the SID within the binding pocket differs (see Figure 7.5a). Although the side chain configuration of residues in the central part of the binding pocket of the PAH domain is largely maintained, subtle rearrangements

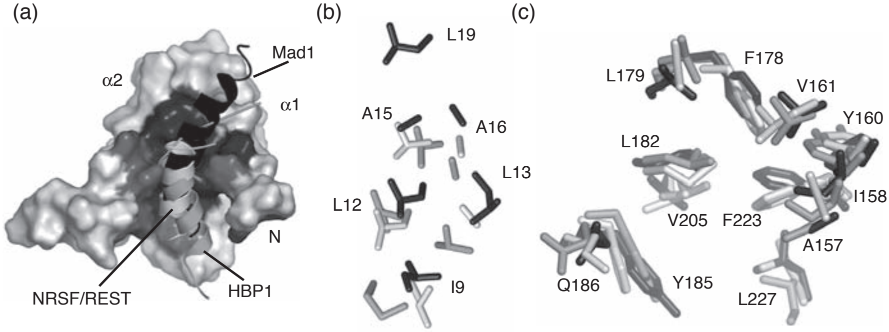


Figure 7.5. Structural adaptivity in the PAH-SID complex. The different PAH-SID complexes and the unbound PAH2 structure are superimposed only on the C_α atoms of the PAH domains. (a) The helices of the SIDs of Mad1 (black), HBP1 (dark grey) and NRSF/REST (light grey) are shown superposed on the structure of the free PAH2 domain. The central part and rim of the binding pocket are indicated as in Figure 7.2. The Mad1-SID helix is slightly bend and longer, allowing residues Y18, L19 and E23 to interact with PAH2. The NRSF/REST-SID is displaced towards helix α2 compared to the Mad1-SID. The HBP1-SID has a relatively disordered tail that almost wraps around helix α1. (b) Close-up on the residues of the consensus-motif in the three repressors (labels refer to the Mad1-SID). The color coding of the SIDs is as in (a). The SIDs are viewed, looking down on the buried side of the SID (rotated 180° with respect to (a) along the helix-axis). Note that the position the side chains of L12 and A15 of Mad1 and the corresponding residues in HBP1 and NRSF/REST are very similar, while there is more variation on the other side of the helix, facing helix α2 of the PAH domain. (c) Close-up of the binding pocket residues in the three complexes, looking along the helix-axis of the SID (rotated 90° with respect to (b) along an axis perpendicular to the paper). Color coding: PAH2 domain in the Mad1 complex in black; PAH2 domain in the HBP1 complex in dark grey; PAH1 domain in the NRSF/REST complex in light grey. Note that the side chain configuration is very similar.

allow for a different packing of the residues of the ϕXXLϕX[V,A]A SID consensus motif, which results in a different position and orientation of the SID helix (Figure 7.5). The angle between the Mad1 and HBP1 helix axis is as large as 35°, the angle between the NRSF/REST and HBP1 helix axis is 20°, and the angle between the Mad1 and NRSF/REST helix axis is 16° In turn, the different position and orientation of the SID allows for unique additional interactions for each repressor, boosting the binding affinity and thus specificity of the interaction.

In conclusion, the consensus motif for the SID defines a basal level of specificity for the PAH-SID interaction. Limited structural arrangements of the binding pocket of the PAH domain allow for an optimized orientation of the SID in which additional interactions, unique for every repressor, are formed. These interactions increase the binding affinity and make the interaction more competitive.

Likely scenario of complex formation

On the basis of the highly similar structures of the PAH2 domain in the unbound and bound state and the presumed transition of the SID from a unfolded state to a helix fold, it is argued in Chapter 5 that the PAH domain acts as a folding template for the SID. To describe the formation of the PAH-SID complex, it useful to discriminate between an encounter complex *AB* and the final complex *C*, where *A* and *B* represent the PAH domain and the SID, respectively:

$$A + B \rightleftarrows AB \rightleftarrows C \qquad [7.1]$$

A reasonable mechanism for complex formation is a combination of induced fit and conformational selection (Ramajani et al., 2004; Smith et al., 2005). In such a model, the PAH domain would bind a subset of conformations of the disordered SID that match the binding pocket to some extent, producing a native-like encounter complex. The formation of the encounter complex might be facilitated by the enhanced flexibility of the N-terminal part of helix α1 in the unbound state. Formation of the encounter complex might also occur via a minor, partially unfolded state of the PAH domain, but, at least for the case of the Mad1-PAH2^B interaction, the dominant pathway will be via the major, folded state, as the binding pocket is perfectly accessible. Once the complex with Mad1 is formed, the equilibrium is shifted such that the concentration of unbound PAH2 becomes very low and exchange between the unbound state and a partially unfolded state can no longer be observed (Chapter 5). Exchange between folded and partially unfolded PAH2 within the complex is most likely negligible, because burial of the binding pocket and the formation of intermolecular interactions will result in a thermodynamic stabilization of the folded state of PAH2.

A few residues of the SID serve as 'anchors' in the initial encounter complex by inserting their side chains into a pocket in the core of the interaction surface of PAH2. I suggest that the two bulky residues in the centre of the consensus interaction motif, L12 and L13 in Mad1, act as such anchor-residues. Subsequently, the SID folds as a helix, guided by the formation of extensive contacts in the binding pocket. As the binding pocket is slightly 'oversized', subtle side chain rearrangements are sufficient to allow optimal positioning of the SID, depending on its chemical character. The binding affinity is further optimized by a induced fit process in the periphery of the binding interface, including the formation of SID-dependent, unique interactions with residues in the rim of the binding pocket.

Suggestions for further research

The research described in this thesis has contributed to a better characterization of the PAH domains and a better understanding of the PAH-SID interaction. Also, this research has generated a number of interesting options for further research. First, I proposed that the PAH3 domain will have the typical four helix bundle fold, whereas the PAH4 domain will likely adopt a distinct fold. Structure determination of the PAH3 and PAH4 domain will prove or disprove this hypothesis. Also, study of the interactions of these domains with their target proteins will be important to understand the function of Sin3.

Secondly, I proposed that the PAH domains form folded modules in full length Sin3 as 'beads-on-a-string'. Characterizing the structure of the short splice variant of mSin3B offers an exciting possibility to study this experimentally. This splice variant comprises roughly 350 residues and should just be amenable to NMR methods. It contains only the PAH1 and PAH2 domain and could give insight regarding the mutual orientation of the PAH domains and regarding the possible presence of structure elements other than the PAH domains.

Furthermore, full length Sin3 contains another highly conserved domain, the HDAC interaction domain. This domain of roughly 280 residues has no homology to other domains with known structure and is crucial for the integrity of the Sin3/HDAC corepressor complex. Most core components of the corepressor complex interact with Sin3 via this domain. The structure of the HID could be used to precisely map the interaction surface with the HDAC proteins, SDS3 and other HID interacting proteins.

Finally, a promising NMR method has been developed that allows characterization of unobservable, transient proteins states that can be very important in protein function. Further development of the CEESY method by its application to nuclei such as the carbonyl and C_α will provide more detailed structural information about these unobservable protein conformations.

References

Abragam, A. (1961). *Principles of Nuclear Magnetism*, Oxford: Clarandon Press.

Adhikary, S. & Eilers, M. (2005). Transcriptional regulation and transformation by MYC proteins. *Nat. Rev. Mol. Cell Biol*,**.** **6**, 635-645.

Ahringer, J. (2000). NuRD and SIN3: histone deacetylase complexes in development. *Trends Genet.*, **16**, 351-356.

Akke, M. (2002). NMR methods for characterizing microsecond to millisecond dynamics in recognition and catalysis. *Curr. Opin. Struct. Biol.*, **12**, 642-647.

Alland, L., Muhle, R., Hou, H. Jr., Potes, J., Chin, L., Schreiber-Agus, N. and DePinho, R.A. (1997). Role for N-CoR and histone deacetylase in Sin3-mediated transcriptional repression, *Nature*, **387**, 49-55.

Alland, L., David, G., Hong, S.L., Potes, J., Muhle, R., Lee, H.C., Hou H., Chen, K. & DePinho, R.A. (2002). Identification of mammalian Sds3 as an integral component of the Sin3/histone deacetylase corepressor complex. *Moll. Cell. Biol.*, **22**, 2743-2750.

Allfrey, V.G., Faulkner, R. and Mirsky, A.E. (1964). Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc. Natl. Acad. Sci. USA*, **51**, 786-794.

Altieri, A.S. and Byrd, R.A. (2004). Automation of NMR structure determination of proteins. *Curr. Opin. Struct. Biol.*, **14**, 547-553.

Aue, W.P., Bartholdi, E. and Ernst, R.R. (1976). Two-dimensional spectroscopy. Application to nuclear magnetic resonance. *J. Chem. Phys.*, **64**, 2229-2246.

Ayer, D.E., Lawrence, Q.A. and Eisenman, R.N. (1995). Mad-Max transcriptional repression is mediated by ternary complex formation with mammalian homologs of yeast repressor Sin3. *Cell*, **80**, 767-776.

Ayer, D.E. (1999). Histone deacetylases: transcriptional repression with SINers and NuRDs. *Trends Cell. Biol.*, **9**, 193- 198.

Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E.L., Studholme, D.J., Yeats, C. and Eddy, S.R. (2004). The Pfam protein families database. *Nucl. Acids Res.*, **32**, D138-41.

Bartels, C.H., Xia, T.H., Billeter, M., Güntert, P. and Wütrich, K. (1995). The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J. Biomol. NMR*, **6**, 1-10.

Bax, A. and Davis, D.G. (1985). MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Magn. Res.*, **65**, 355-360.

Bax, A. and Grishaev, A. (2005). Weak alignment NMR: a hawk-eyed view of biomolecular structure. *Curr. Opin. Struct. Biol.*, **15**, 563-570.

Beach, H., Cole, R., Gill, M.L. & Loria, J.P. (2005). Conservation of mus-ms enzyme motions in the apo- and substrate-mimicked state. *J. Am. Chem. Soc.*, **127**, 9167-9176.

Berger, S.L. (2002). Histone modifications in transcriptional regulation, *Curr. Opin. Genet. Dev.*, **12**, 142-148.

Berger, S.L. (1999). Gene activation by histone and factor acetyltransferases. *Curr. Opin. Cell Biol.*, **11**, 336-341.

Berman, H.M., Henrick, K. and Nakamura, H. (2003). Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol.*, **10**, 980.

Billeter, M., Ian, Y.Q., Otting, G., Muller, M., Gering, W. and Wüthrich K. (1993). Determination of the Nuclear Magnetic Resonance solution structure of an Antennapedia homeodomain-DNA complex. *J. Mol. Biol.*, **234**, 1084-1094.

Bloch, F., Hansen, W.W. and Packard, M. (1946). Nuclear induction. *Phys. Rev.*, **69**, 127.

Braunschweiler, L. and Ernst, R.R. (1983). Coherence transfer by isotropic mixing: application to proton correlation spectroscopy. *J. Magn. Res.*, **53**, 521-528.

Brubaker, K., Cowley, S.M., Huang, K., Loo, V., Yochum, G.S., Ayer, D.E., Eisenman, R.N. and Radhakrishnan, I. (2000). Solution structure of the interacting domains of the Mad-Sin3 complex: implications for recruitment of a chromatin-modifying complex. *Cell*, **103**, 655-665.

Brünger, A.T. (1996). X-PLOR Version 3.851. *A system for X-ray crystallography and NMR*, Yale University Press, New Haven.

Burke, L.J. & Baniahmad, A. (2000). Co-repressors 2000. *FASEB J.*, **14**, 1876-1888.

Carver, J.P & Richards, R.E. (1972). A general two-site solution for the chemical exchange produced dependence of T2 upon the Carr-Purcell pulse separation. *J. Magn Res.*, **6**, 89-105.

Case, D.A. (1998). The use of chemical shifts and their anisotropies in biomolecular structure determination. *Curr. Opin. Struct. Biol.*, **8**, 624-630.

Cavanagh, J. and Rance, M. (1990). Sensitivity improvement in isotropic mixing (TOCSY) experiments. *J. Magn. Reson.*, **88**, 72-85.

Cavanagh, J., Fairbrother, W.J., Palmer, A.G. and Skelton, N.J. (1996). *Protein NMR spectroscopy: principles and practice.* San Diego: Academic Press.

Chuprina, V.P., Rullmann, J.A.C., Lamerichs, R.M.J.N., Boom, J.H. van, Boelens, R. and Kaptein, R. (1994). Structure of the complex of Lac repressor headpiece and an 11 base-pair half-operator determined by Nuclear Magnetic Resonance spectroscopy and restrained Molecular Dynamics. *J. Mol. Biol.*, **234**, 446-462.

Cieslar, C., Holak, T.A. and Oschkinat, H. (1990). 3D TOCSY-TOCSY Processing using linear prediction, as a potential technique for automated assignment. *J. Magn. Reson.*, **89**, 184-190.

Clore, G.M., Szabo, A., Bax, A., Kay, L.E., Driscoll, P.C. and Gronenborn, A.M. (1990). Deviations from the simple 2-parameter model-free approach to the interpretation of N-15 nuclear magnetic-relaxation of proteins. *J. Am. Chem. Soc.*, **112**, 4989-4991.

Clore, G.,M. and Gronenborn, A.M. (2005). New methods of structure refinement for macromolecular structure determination by NMR. *Proc. Natl. Acad. Sci. USA*, **95**, 5891-5898.

Congreve, M., Murray, C.W. and Blundell, T.L. (2005). Structural biology and drug discovery. *Drug Discov. Today*, **10**, 895-907.

Cornilescu G., Delaglio F. & Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology *J. Biomol. NMR*, **13**, 289-302.

Cosma, M.P. (2002). Ordered recruitment: gene-specific mechanism of transcription activation. *Mol. Cell*, **10**, 227-238.

Cowley, S.M., Kang, R.S., Frangioni, J.V., Yada, J.J., DeGrand, A.M., Radhakrishnan, I. & Eisenman, R.N. (2004). Functional analysis of the Mad1-mSin3A repressor-corepressor interaction reveals determinants of specificity, affinity, and transcriptional response. *Mol. Cell. Biol.*, **24**, 2698-2709.

Cowley, S.M., Iritani, B.M., Mendrysa, S.M., Xu, T., Cheng, P.F., Yada, J., Ligitt, H.D. & Eisenman, R.N. (2005) The mSin3A chromatin-modifying complex is essential for embryogenesis and T-cell development. *Mol. Cell. Biol.*, **25**, 6990-7004.

Dang, V.D., Benedik, M.J., Ekwall, K., Choi, J., Allshire, R.C., Levin, H.L. (1999). A new member of the Sin3 family of corepressors is essential for cell viability and required for retroelement propagation in fission yeast, *Mol. Cell. Biol.*, **19**, 2351-2365.

Dannenberg, J.-H., David, G., Zhong, S., Torre, J., Wong, W.H. & DePinho, R.A. (2005). mSin3A corepressor regulates diverse transcriptional networks governing normal and neoplastic growth and survival. *Genes Dev.*, **19**, 1581-1592.

David, G., Alland, L., Hong, S.H., Wong, C.W., DePinho, R.A. and Dejean, A. (1998). Histone deacetylase associated with mSin3A mediates repression by the acute promyelocytic leukemia-associated PLZF protein. *Oncogene*, **16**, 2549-2556.

David, G., Turner, G.M., Yao, Y., Protopopov, A. and DePinho, R.A. (2003). mSin3-associated protein, mSds3, is essential for pericentric heterochromatin formation and chromosome segregation in mammalian cells, *Genes Dev.*, **17**, 2396-2405.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeiffer, J. and Bax, A. (1995). NMRPipe – a multidimensional processing system based on UNIX pipes, *J. Biomol. NMR*, **4**, 277-293.

Dijkstra, R.W. (1989). A method to suppress cycling sidebands in broadband decoupling, *J. Magn. Reson.*, **82**, 347-351.

Dongen, M. van, Weigelt, J., Uppenberg, J., Schultz, J. and Wikstrom, M. (2002). Structure-based screening and design in drug discovery. *Drug Discov. Today*, **7**, 471-478.

Dosset, P., Hus, J.C., Blackledge, M. and Marion, D. (2000). Efficient analysis of Macromolecular Rotational Diffusion from Heteronuclear Relaxation Data. *J. Biomol. NMR*, **16**, 23-28.

Dyson, H.J. and Wright, P.E. (2005). Elucidation of the protein folding landscape by NMR. *Methods Enzymol.*, **394**, 299-321.

Eberharter, A. & Becker, P.B. (2002). Histone acetylation: a switch between repressive and permissive chromatin. Second in review series on chromatin dynamics. *EMBO Rep.*, **3**, 224-229.

Eilers, A.L., Billin, A.N., Liu, J. and Ayers, D.E. (1999). A 13-amino acid amphipathic α-helix is required for the functional interaction between the transcriptional repressor Mad1 and mSin3A. *J. Biol. Chem.*, **274**, 32750-32756.

Eisenmesser, E. Z., Bosco, D. A., Akke, M. and Kern, D. (2002). Enzyme dynamics during catalysis. *Science*, **295**, 1520-1523.

Eissenmesser, E.Z., Millet, O., Labeikovsky, W., Korzhnev, D.M., Wolf-Watz, M., Bosco, D.A., Skalicky, J.J, Kay, L.E., Kern, D. (2005). Intrinsic dynamics of an enzyme underlies catalysis. *Nature*, **438**, 117-121.

Engelke, J. and Rüterjans, H. (1999). Recent developments in studying the dynamics of protein structures from ^{15}N and ^{13}C relaxation time measurements. In: *Biological Magnetic Resonance vol. 17: Structure Computation and Dynamics in Protein NMR* (Krishna, N.R. and Berliner, L.J. Eds.), Kluwer Academic/Plenum Publishers, New York.

Ernst, R.R., Bodenhausen, G. and Wokaun, A. (1987). *Principles of nuclear magnetic resonance in one and two dimensions*, Oxford: Clarendon Press.

Farrow, N.A, Zhang, O., Szabo, A., Torchia, D.A. & Kay, L.E. (1995). Spectral density function mapping using ^{15}N relaxation data exclusively. *J. Biomol. NMR*, **6**, 153-162.

Featherstone, M. (2002). Coactivators in transcription initiation: here are your orders. *Curr. Opin. Genet. Dev.*, **12**, 149-155.

Felsenfeld, G. and Groudline, M. (2003). Controlling the double helix. *Nature*, **421**, 448-453.

Feng, Q. and Zhang, Y. (2003). The NuRD complex: linking histone modification to nucleosome remodeling. *Curr. Top. Microbiol. Immunol.*, **274**, 269-290.

Fischer, E. (1894). Einfluss der Configuration auf die Wirkung der Enzyme. *Ber. Dtsch. Chem.*

Ges., **27**, 2984-2993.

Fischle, W., Wang, Y. and Allis, C.D. (2003). Histone and chromatin cross-talk. *Curr. Opin. Cell Biol.*, **15**, 172-183.

Fleischer, T.C., Yun, U.J. & Ayer, D.E. (2003). Identification and characterization of three new components of the mSin3A Corepressor complex. *Mol. Cell. Biol.*, **23**, 3456-3467.

Foley, K.P. and Eisenman, R.N. (1999). Two MAD tails: what the recent knockouts of Mad1 and Mxi1 tell us about the MYC/MAX/MAD network. *Biochim. Biophys. Acta*, **1423**, M37-M47.

Frauenfelder, H. and Gratton, E. (1986). Protein dynamics and hydration, *Meth. Enzym.*, **127**, 207-216.

Freeman, R., Frenkiel, T. and Levitt, M.H. (1982). A simple "black-box" decoupler, *J. Magn. Reson.*, **50**, 345-348.

Grandori, C., Cowley, S.M., James, L.P. and Eisenman, R.N. (2000). The Myc/Max/Mad network of the transcriptional control of cell behavior. *Ann. Rev. Cell. Dev. Biol.*, **16**, 653-699.

Grey, M.J., Wang, C. & Palmer, A.G. (2003). Disulfide bond isomerization in basic pancreatic trypsin inhibitor: multisite chemical exchange quantified by CPMG relaxation dispersion and chemical shift modeling. *J. Am. Chem. Soc.*, **125**, 14324-14335.

Grunstein, M. (1997). Histone acetylation in chromatin structure and transcription. *Nature*, **389**, 349-352.

le Guezennec, X., Vriend, G. and Stunnenberg, H.G. (2004). Molecular determinants of the interaction of Mad with the PAH2 domain of mSin3. *J.Biol. Chem.*, **279**, 25823-25829.

Güntert, P., Mumenthaler, C. & Wüthrich, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.*, **273**, 283-298.

Gurd, F.R.N. and Rothgeb, T.M. (1979). Motions in proteins. *Adv. Protein Chem.*, **33**, 73-165.

Halleck, M.S., Pownall, S., Harder, K.W., Duncan, A.M.V., Jirik, F.R. & Schlegel, R.A. (1995). A widely distributed putative mammalian transcriptional regulator containing multiple paired amphipathic helices, with similarity to yeast SIN3. *Genomics*, **26**, 403-406.

Hassig, C.A., Fleischer, T.C., Billin, A.N., Schreiber, S.L. & Ayer, D.E. (1997). Histone deacetylase activity is required for full transcriptional repression by mSin3A. *Cell*, **89**, 341-347.

Heinzel, T., Lavinsky, R.M., Mullen, T.M., Soderstrom, M., Laherty, C.D., Torchia, J., Yang, W.M., Brard, G., Ngo, S.D., Davie, J.R., Seto, E., Eisenman, R.N., Rose, D.W., Glass, C.K. and Rosenfeld, M.G. (1997). A complex containing N-CoR, mSin3 and histone deacetylase mediates transcriptional repression. *Nature*, **387**, 43-48.

Herrmann, T., Güntert, P. & Wüthrich, K. (2002). Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.*, **319**, 209-227.

Hooft, R.W.W., Vriend, G., Sander, C. & Abola, E.E. (1996). Errors in protein structures. *Nature*, **381**, 272-272.

Hu, X. and Lazar, M.A. (2000). Transcriptional repression by nuclear hormone receptors. *Trends Endocrinol. Metab.*, **11**, 6-10.

Ikura, M. and Bax, A. (1992). Isotope-filtered 2D NMR of a protein peptide complex – Study of a skeletal-muscle myosin light chain kinase fragment bound to calmodulin. *J. Am. Chem. Soc.*, **114**, 2433-2440.

Ikura, M., Clore, G.M., Gronenborn, A.M., Zhu., G., Klee, C.B. and Bax, A. (1992). Solution

Structure of a calmodulin-target peptide complex by multidimensional NMR. *Science*, **256**, 632-638.

van Ingen, H., Tessari, M. and Vuister, G.W. (2002). A 3D doubly sensitivity enhanced X-filtered TOCSY-TOCSY experiment. *J. Biomol. NMR*, **24**, 155-160.

van Ingen, H., Lasonder, E., Jansen, J.F.A., Kaan, A.M., Spronk, C.A.E.M., Stunnenberg, H.G.M. & Vuister, G.W. (2004). Extension of the binding motif of the Sin3 interacting domain of the Mad family proteins. *Biochem.*, **43**, 46-54.

van Ingen, H., Baltussen, M.A.H., Aelen, J.M.A, Vuister, G.W. (2006). Role of structural and dynamical plasticity in Sin3: the free PAH2 domain is a folded module in mSin3B. *J. Mol. Biol.* , **358.**, 485-497.

van Ingen, H., Vuister, G.W., Wijmega, S & Tessari, M. (2006). CEESY: Characterizing the conformation of unobservable protein states. *J. Am. Chem. Soc.*, **128**, 3856-3857.

Ishima, R. & Nagayama, K. (1996). Quasi-spectral-density function analysis for nitrogen-15 nuclei in proteins. *J. Magn. Res. Ser. B.*, **108**, 73-76.

Ishima, R., Torchia, D.A. (2003). Extending the range of amide proton relaxation dispersion experiments in proteins using a constant-time relaxation-compensated CPMG approach. *J. Biomol. NMR*, **25**, 243-248.

Ishima, R, Baber, J., Louis, J.M., Torchia, D.A. (2004). Carbonyl carbon transverse relaxation dispersion measurements and ms-micros timescale motion in a protein hydrogen bond network. *J. Biomol. NMR*, **29**, 187-198.

Jeener, J., Meier, B. H., Bachmann, P. and Ernst, R. R. (1979). Investigations of exchange processes by two-dimensional NMR spectroscopy. *J. Chem. Phys.*, **69**, 4546-4553.

Jen, J. (1978). Chemical exchange and NMR T2 relaxation – the multisite case. *J. Magn. Res.*, **30**, 111-128.

Jones, P.L., Veenstra, G.J., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, J. and Wolffe, A.P. (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat. Genet.*, **19**, 187-191.

Juwelein, T. and Allis, C.D. (2001). Translating the histone code. *Science*, **293**, 1074-1080.

Kadosh, D. and Struhl, K. (1997). Repression by Ume6 involves recruitment of a complex containing Sin3 corepressor and Rpd3 histone deacetylase to target promoters. *Cell*, **89**, 365-371.

Kalkhoven, E. (2004). CBP and p300: HATs for different occasions. *Biochem. Pharmacol.*, **68**, 1145-1155.

Karplus, M. (1963). Vicinal proton coupling in Nuclear Magnetic Resonance. *J. Am. Chem. Soc.*, **85**, 2870-2871.

Kay, L.E., Keifer, P. and Saarinen, T. (1992). Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.*, **114**, 10663-10665.

Kay, L.E. (1998). Protein dynamics from NMR. *Biochem. Cell Biol.*, **76**, 142-152.

Kern, D. and Zuiderweg, E.R.P. (2003). The role of dynamics in allosteric regulation. *Curr. Opin. Struct. Biol.*, **13**, 748-757.

Khan, M.M., Nomura, T., Kim, H., Kaul, S.C., Wadhwa, R., Shinagawa, T., Ichikawa-Iwata, E., Zhong, S., Pandolfi, P.P., Ishii, S. (2001). Role of PML and PML-RARalpha in Mad-mediated transcriptional repression, *Mol. Cell*, **7**, 1233-1243.

Kim, H., Lee, J.-E., Cho, E.-J., Lio, J.O. and Youn, H.-D. (2003). Menin, a tumor supressor, represses JunD-mediated transcriptional activity by association with an mSin3A-histone deacetylase complex, *Cancer Res.*, **63**, 6135-6139.

Koradi, R., Billeter, M., and Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics*, **14**, 51-55.

Kornberg, R.D. and Lorch, Y. (1999). Twenty-five years of the nucleosome, fundamental particle of the eukaryotic chromosome, *Cell*, **98**, 285-294.

Kornberg, R.D. (2005). Mediator and the mechanism of transcriptional activation. *Trends Biochem. Sci.*, **30**, 235-239.

Korzhnev, D.M., Orekhov, V.Y., Dahlquist, F.W., Kay, L.E. (2003). Off-resonance R1rho relaxation outside of the fast exchange limit: an experimental study of a cavity mutant of T4 lysozyme. *J. Biomol. NMR*, **26**, 39-48.

Korzhnev, D.M., Salvatella, X., Vendruscolo, M., Di Nardo, A.A., Davidson, A.R., Dobson, C.M. and Kay, L.E. (2004). Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature*, **430**, 586-590.

Korzhnev, D.M., Orekhov, V.Y., Kay, L.E. (2005). Off-resonance R1rho NMR studies of exchange dynamics in proteins with low spin-lock fields: an application to a Fyn SH3 domain. *J. Am. Chem. Soc.*, **127**, 713-721.

Koshland, D.E. Jr. (1958). Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl. Acad. Sci. USA*, **44**, 98-104.

Köver, K.E., Uhrín, D. and Hruby, V.J. (1998). Gradient- and sensitivity-enhanced TOCSY experiments. *J. Magn. Reson.*, **130**, 162-168.

Krishnamurthy, V.V. (1995). Sensitivity-enhanced 3D HSQC-TOCSY experiments. *J. Magn. Reson. Ser. B*, **106**, 170-177.

Kumar, A., Ernst, R.R. and Wüthrich, K. (1980). A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochem. Biophys. Res. Comm.*, **95**, 1-6.

Kumar, S. and Nussinov, R. (2002). Relationship between ion pair geometries and electrostatic strengths in proteins. *Biophys. J.*, **83**, 1595-1612.

Kupče, E., Freeman, R., Wider, G. and Wüthrich, K. (1996). Suppression of cycling sidebands using bi-level adiabatic decoupling, *J. Magn. Reson. Ser. A*, **122**, 81-84.

Kupče, E. and Freeman, R. (2001). A two-dimensional experiment that separated decoupling sidebands from the main peaks, *J. Magn. Reson.*, **151**, 142-145.

Kuzmichev, A., Zhang, Y., Erdjument-Bromage, H., Tempst, P. and Reinberg, D. (2002). Role of the Sin3-histone deacetylase complex in growth regulation by the candidate tumor suppressor p33ING1. *Mol. Cell. Biol.*, **22**, 835-848.

Laherty, C.D., Yang, W.M., Sun, J.M., Davie, J.R., Seto, E. & Eisenman, R.N. (1997). Histone deacetylases associated with the Sin3 corepressor mediate Mad transcriptional repression. *Cell*, **89**, 349-356.

Lai, A., Kennedy, B.K., Barbie, D.A., Bertos, N.R., Yang, X.J., Theberge, M.C., Tsai, S.C., Seto, E., Zhang, Y., Kuzmichev, A., Lane, W.S., Reinberg, D., Harlow, E. and Branton, P.E. (2001). RBP1 recruits the mSIN3-histone deacetylase complex to the pocket of retinoblastoma tumor suppressor family proteins found in limited discrete regions of the nucleus at growth arrest, *Mol. Cell. Biol.*, **21**, 2918-2932.

Laskowski, R.A., Rullmann, J.A., MacArthur, M.W., Kaptein, R. and Thornton, J.M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR*, **8**, 477-486.

Lefèvre, J.F., Dayie, K.T., Peng, J.W. & Wagner, G. (1996). Internal mobility in the partially folded DNA binding and dimerization domains of GAL4: NMR analysis of the N-H spectral density functions. *Biochem.*, **35**, 2674-2686.

Levitt, M.H. and Freeman, R. (1981). Composite pulse decoupling. *J. Magn. Res.*, **43**, 502-507.

Levitt, M.H., Freeman, R. and Frenkiel, T. (1982). Broadband heteronuclear decoupling, *J. Magn. Reson.*, **47**, 328-330.

Levitt, M.H. (1997). The signs of frequencies and phases in NMR. *J. Magn. Reson.*, **126**, 164-182.

Levitt, M.H., Johannessen, O.G. (2000). Signs of frequencies and phases in NMR: the role of radiofrequency mixing. *J. Magn. Reson.*, **142**, 190-194.

Levitt, M.H. (2001). *Spin dynamics: basics of nuclear magnetic resonance*. Chichester: John Wiley & Sons, Ltd.

Li, X., Keskin, O., Ma, B., Nussinov, R. & Liang J. (2004). Protein-protein interactions: hot spots and structurally conserved residues often locate in complemented pockets that pre-organized in the unbound states: implications for docking. *J. Mol. Biol.*, **344**, 781-795.

Lin, R.J., Nagy, L., Inoue, S., Shao, W., Miller, W.H. Jr. and Evans, R.M. (1998). Role of the histone deacetylase complex in acute promyelocytic leukaemia, *Nature*, **391**, 811-814.

Linge, J.P., Williams, M.A., Spronk, C.A.E.M, Bonvin, A.J.J. and Nilges, M. (2003). Refinement of protein structures in explicit solvent. *Proteins*, **50**, 496-506.

Lipari, G. & Szabo, A. (1982a). Model-free approach to the interpretation of nuclear magnetic resonance in macromolecules 1: Theory and range of validity. *J. Am. Chem. Soc.*, **104**, 4546-4559.

Lipari, G. & Szabo, A. (1982b). Model-free approach to the interpretation of nuclear magnetic resonance in macromolecules 2: Analysis of experimental results. *J. Am. Chem. Soc.*, **104**, 4559-4570.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F. and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, **389**, 251-260.

Lundström, P., Akke, M. (2005). Microsecond protein dynamics measured by ^{13}Calpha rotating-frame spin relaxation. *ChemBioChem*, **6**, 1685-1692.

Lüscher, B. (2001). Function and regulation of the transcription factors of the Myc/Max/Mad network. *Gene*, **277**, 1-14.

Lo Conte, L., Chothia, C. and Janin, J. (1999). The atomic structure of protein-protein recognition sites. *J. Mol. Biol.*, **285**, 2177-2198.

Logie, C., Tse, C., Hansen, J.C. and Peterson, C.L. (1999). The core histone N-terminal domains are required for multiple rounds of catalytic chromatin remodeling by the SWI/SNF and RSC complexes. *Biochem.*, **38**, 2514-2522.

Lorch, Y., Beve, J., Gustafsson, C.M, Myers, L.C and Kornberg, R.D. (2000). Mediator-Nucleosome interactions. *Mol. Cell*, **6**, 197-201.

Loria, J.P., Rance, M. & Palmer, A.G. (1999) A relaxation-compensated Carr-Purcell-Meiboom-Gill sequence for characterizing chemical exchange by NMR spectroscopy. *J. Am. Chem. Soc.*, **121**, 2331-2332.

Mandel, A.M., Akke, M. and Palmer, A.G. (1995). Backbone dynamics of Escherichia Coli Ribonuclease HI: correlations with structure and function in an active enzyme. *J. Mol. Biol.*, **246**, 144-163.

Marchler-Bauer, A. & Bryant, S.H. (2004). CD-search: protein domain annotations on the fly. *Nucl. Acids. Res.*, **32**, W327-331.

Marion, D., Ikura, M., Tschudin, R. and Bax, A. (1989). Rapid recording of 2D NMR spectra without phase cycling. Application to the study of hydrogen exchange in proteins, *J. Magn. Reson.*, **85**, 393-399.

McConnell, H.M. (1958). Reaction rates by nuclear magnetic resonance. *J. Chem. Phys.*, **28**, 430-431.

McElheny, D., Schnell, J.R., Lansing, J.C., Dyson, H.J. & Wright, P.E. (2005). Defining the role of active-site loop fluctuations in dihydrofolate reductase catalysis. *Proc. Natl. Acad. Sci. USA*, **102**, 5032-5037.

Mellor, J. (2005). The dynamics of chromatin remodelling at promoters. *Mol. Cell*, **19**, 147-157.

Melnick, A. and Licht, J.D. (2002). Histone deacetylases as therapeutic targets in hematologic malignancies, *Curr. Opin. Hematology*, **22**, 322-332.

Mulder F.A., Skrynnikov N.R., Hon B., Dahlquist F.W. & Kay L.E. (2001) Measurement of slow (micros-ms) time scale dynamics in protein side chains by ^{15}N relaxation dispersion NMR spectroscopy: application to Asn and Gln residues in a cavity mutant of T4 lysozyme. *J. Am. Chem. Soc.*, **123**, 967-975.

Mulder, F.A., Mittermaier, A., Hon, B., Dahlquist, F.W. and Kay, L.E. (2001). Studying excited states of proteins by NMR spectroscopy. *Nat. Struct. Biol.*, **8**, 932-935.

Murphy, M., Ahn, J., Walker, K.K., Hoffman, W.H., Evans, R.M., Levine, A.J. and George, D.L. (1999). Transcriptional repression by wild-type p53 utilizes histone deacetylases, mediated by interaction with mSin3A. *Genes Dev.*, **13**, 2490-2501.

Nabuurs, S.B., Krieger, E., Spronk, C.A.E.M, Nederveen, A.J., Vriend, G. and Vuister, G.W. (2005). Definition of a new information based per residue quality parameter. *J. Biomol. NMR*, **33**, 123-134.

Nadassy, K., Wodak, S.J. and Janin, J. (1999). Structural features of protein-nucleic acid recognition sites. *Biochem.*, **38**, 1999-2017.

Nagy, L. Kao, H.-Y., Chakravarti, D., Lin, R.J., Hassig, C.A., Ayer, D.E., Schreiber, S.L. and Evans, R.M. (1997). Nuclear receptor repression mediated by a complex containing SMRT, mSin3A, and histone deacetylase, *Cell*, **89**, 373-380.

Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N. and Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature*, **393**, 386-389.

Neri, D., Szyperski, T., Otting, G., Senn, H. & Wüthrich, K. (1989) Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional ^{13}C labeling. *Biochem.*, **28**, 7510-7516.

Nomura, M., Uda-Tochio, H., Muurai, K., Mori, N. and Nishimura, Y. (2005). The neural repressor NRSF/REST binds the PAH1 domain of the Sin3 corepressor by using its distinct short hydrophobic helix. *J. Mol. Biol.*, **354**, 903-915.

Omichinski, J.G., Clore, G.M., Schaad, O., Felsenfeld, G., Trainor, C., Apella, E., Stahl, S.J. and Gronenborn, A.M. (1993). NMR structure of a specific DNA complex of Zn-containing DNA-binding domain of GATA-1. *Science*, **261**, 438-446.

Orphinades, G. and Reinberg, D. (2002). A unified theory of gene expression. *Cell*, **108**, 439-451.

Otting, G., Senn, H., Wagner, G. and Wüthrich, K. (1986). Editing of 2D ^{1}H NMR spectra using X half-filters. Combined use with residue-selective ^{15}N-labeling of proteins. *J. Magn. Reson.*, **70**, 500-505.

Palmer, A.G., Rance, M. and Wright, P.E. (1991). Intramolecular motions of a Zn finger DNA-binding domain from Xfin characterized by proton-detected natural abundance ^{13}C heteronuclear NMR spectroscopy. *J. Am. Chem. Soc.*, **113**, 4371-4380.

Palmer, A.G. (1997). Probing molecular motion by NMR, *Curr. Opin. Struct. Biol.*, **7**, 732-737.

Palmer, A.G., Kroenke, C.D. and Loria, J.P. (2001). Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol.*, **339**, 204-238.

Palmer, A.G., Grey, M.J., Wang, C. (2005). Solution NMR spin relaxation methods for characterizing chemical exchange in high-molecular-weight systems. *Methods Enzymol.*, **394**, 430-465.

Pang, Y.-P., Kumar, G.A., Zhang, J.-S. and Urrutia, R. (2003). Differential binding of Sin3 interacting repressor domains to the PAH2 domain of Sin3A. *FEBS Lett.*, **548**, 108-112.

Piotto, M., Saudek, V. and Sklenar, V. (1992). Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR*, **2**, 661-665.

Pochapsky, S.S. and Pochapsky, T.C. (2001). Nuclear magnetic resonance as a tool in drug discovery, metabolism and disposition. *Curr. Top. Med. Chem.*, **1**, 427-441.

Ptashne, M. & Gann, A. (1997). Transcriptional activation by recruitment. *Nature*, **386**, 569-577.

Ptashne, M. (2005). Regulation of transcription: from lambda to eukaryotes. *Trends Biochem. Sci.*, **30**, 275-278.

Purcell, E.M., Torrey, H.C. and Pound, R.V. (1946). Resonance absorption by nuclear magnetic moments in a solid. *Phys. Rev.*, **69**, 37-38.

Qin, J., Vinograda, O. and Gronenborn, A.M. (2001). Protein-protein interactions probed by nuclear magnetic resonance spectroscopy. *Meth. Enzymology*, **339**, 377-389.

Ramajani, D., Thiel, S., Vajda, S. & Camacho, C.J. (2004). Anchor residues in protein-protein interactions. *Proc. Natl. Acad. Sci. USA*, **101**, 11287-11292.

Rampalli, S., Pavithra, L., Bhatt, A., Kundu, T.K. and Chattopadhyay (2005). Tumor suppressor SMAR1 mediates Cyclin D1 repression by recruitment of the SIN3/histone deacetylase 1 complex, *Mol. Cell. Biol.*, **25**, 8415-8429.

Rayman, J.B., Takahashi, Y., Indjeian, V.B., Dannenberg, J.H., Catchpole, S., Watson, R.J., te Riele, H. and Dynlacht, B.D. (2002). E2F mediates cell cycle-dependent transcriptional repression *in vivo* by recruitment of an HDAC1/mSin3B corepressor complex, *Genes Dev.*, **16**, 933-947.

Roeder, R.G. (2003). The eukaryotic transcriptional machinery: complexities and mechanisms unforeseen. *Nat. Med.*, **9**, 1239-1244.

Roeder, R.G. (2005). Transcriptional regulation and the role of diverse coactivators in animal cells. *FEBS Lett.*, **579**, 909-915.

Roehrl, M.H., Heffron, G.J., Wagner, G. (2005). Correspondence between spin-dynamic phases and pulse program phases of NMR spectrometers. *J. Magn. Reson.*, **174**, 325-330.

Rost, B. (1999). Twilight zone of protein sequence alignments. *Protein Eng.*, **12**, 85-94.

Rutherford, T.J. and Homans, S.W. (1995). Proton resonance assignments in oligosaccharides containing multiple monosaccharide residues of the same type. *J. Magn. Reson. Ser. B*, **106**, 10-13.

Sattler, M., Schwendiger, M.G., Schleucher, J. and Griesinger, C. (1995). Novel strategies for sensitivity enhancement in heteronuclear multidimensional NMR experiments employing pulsed field gradients. *J. Biomol. NMR*, **5**, 11-22.

Schreiber-Agus, N., Chin, L., Chen, K., Torres, R., Rao, G., Guida, P., Skoultchi, A.I. and DePinho, R.A. (1995). An amino-terminal domain of Mxi1 mediates anti-Myc oncogenic activity and interacts with a homolog of the yeast transcriptional repressor SIN3. *Cell*, **80**, 777-786.

Schreiber-Agus, N. and DePinho, R.A. (1998). Repression by the Mad(Mxi1)-Sin3 complex. *Bioessays*, **20**, 808-818.

Schwarzinger, S., Kroon, G. J. A., Foss, T.R., Wright, P. E., Dyson, H. J. (2000). Random coil chemical shifts in acidic 8 M urea: implementation of random coil shift data in NMRView. *J. Biomol. NMR*, **18**, 43-48.

Shaka, A.J., Keeler, J., Frenkiel, T. and Freeman R. (1983a). An improved sequence for broadband decoupling: WALTZ-16, *J. Magn. Reson.*, **52**, 335-338.

Shaka, A.J., Keeler, J. and Freeman, R. (1983b). Evaluation of a new broadband decoupling sequence: WALTZ-16, *J. Magn. Reson.*, **53**, 313-340.

Shaka, A.J., Barker, P.B. and Freeman, R. (1985). Computer-optimized decoupling scheme for wideband applications and low-level operation, *J. Magn. Reson.*, **64**, 547-552.

Shaka, A.J., Barker, P.B., Bauer, C.J. and Freeman, R. (1986). Cycling sidebands in broadband decoupling, *J. Magn. Reson.*, **67**, 396-401.

Shaka, A.J., Lee, C.J. and Pines, A. (1988). Iterative schemes for bilinear operators; application to spin decoupling. *J. Magn. Reson.*, **77**, 274-293.

Silverstein, R.A., Richardson, W., Levin, H., Allshire, R. and Ekwall, K. (2003). A new role for the transcriptional corepressor SIN3; regulation of centromeres, *Curr. Biol.*, **13**, 68-72.

Silverstein, R.A. and Ekwall, K. (2005). Sin3: a flexible regulator of global gene expression and genome stability, *Curr. Genet.*, **47**, 1-17.

Sklenar, V., Piotto, M., Leppik, R. and Saudek, V. (1993). Gradient-tailored water suppression for ^{1}H-^{15}N HSQC experiments optimized to retain full sensitivity. *J. Magn. Reson. Ser. A*, **102**, 241-245.

Skrynnikov, N.R., Dahlquist, F.W. & Kay, L.E. (2002). Reconstructing NMR spectra of "invisible" excited protein states using HSQC and HMQC experiments. *J. Am. Chem. Soc.*, **124**, 12352-12360.

Smith, G., Sternberg, M.J.E. & Bates, P.A. (2005). The relationship between the flexibility of proteins and their conformational states on forming protein-protein complexes with an application to protein-protein docking. *J. Mol.Biol.*, **347**, 1077-1101.

Snyder, D.A., Chen, Y., Denissova, N.G., Acton, T., Aramini, J.M., Ciano, M., Karlin, R., Liu, J., Manor, P., Rajan, P.A., Rossi, P., Swapna, G.V.T., Xiao, R., Rost, B., Hunt, J. and Montelione, G.T. (2005). Comparisons of NMR spectral quality and success in crystallization demonstrate that NMR and X-ray crystallography are complementary methods for small protein structure determination. *J. Am. Chem. Soc.*, **127**, 16505-16511.

Spronk, C.A.E.M., Tessari, M., Kaan, A.M., Jansen, J.F.A., Vermeulen, M., Stunnenberg, H.G. and Vuister, G.W. (2000). The Mad1-Sin3B interaction involves a novel helical fold. *Nat. Struct. Biol.*, **7**, 1100-1104.

Spronk, C.A.E.M., Jansen, J.F.A., Tessari, M., Kaan, A.M., Aelen, J., Lasonder, E., Stunnenberg, H.G. and Vuister, G.W. (2001). Letter to the editor: sequence-specific assignment of the PAH2 domain of Sin3B free and bound to Mad1. *J. Biomol. NMR*, **19**, 377-378.

Spronk, C.A.E.M, Linge, J.P., Hilbers, C.W. and Vuister, G.W. (2002). Improving the quality of protein structures derived by NMR spectroscopy. *J. Biomol. NMR*, **22**, 281-289.

Spronk, C.A.E.M, Nabuurs, S.B, Krieger, E., Vriend, G. and Vuister, G.W. (2004). Validation of protein structures derived by NMR spectroscopy. *Progr. NMR. Spectr.*, **45**, 315-337.

Sternberg, P.W., Stern, M.J., Clark, I. and Herskowitz, I. (1987). Activation of the yeast HO gene by release from multiple negative control. *Cell*, **48**, 567-577.

Struhl, K. (1998). Histone acetylation and transcriptional regulatory mechanisms. *Genes Dev.*, **12**, 599-606.

Sudarsanam, P. and Winston, F. (2000). The Swi/Snf family: nucleosome remodeling complexes and transcriptional control. *Trends Genet.*, **16**, 345-350.

Swanson, K.A, Knoepfler, P.S., Huang, K., Kang, R.S, Cowley, S.M., Laherty, C.D., Eisenman, R.N. & Radhakrishnan, I. (2004). HBP1 and Mad1 repressors bind the Sin3 corepressor PAH2 domain with opposite helical orientations. *Nat. Struct. Mol. Biol.*, **11**, 738-746.

Tanikawa, J., Nomura, T., Macmillan, E.M., Shinagawa, T., Jin, W., Kokura, K., Baba, D., Shirakawa, M., Gonda, T.J. and Ishii, S. (2004). p53 suppresses c-Myb-induced transactivation and transformation by recruiting the corepressor mSin3A, *J. Biol. Chem.*, **279**, 55393-55400.

Theriault, Y., Logan, T.M., Meadows., R., Yu, L.P., Olejniczak, E.T., Holzman, T.F., Simmer, R.L. and Fesik, S.W. (1993). Solution structure of the cyclosporine-A cyclophilin complex by NMR. *Nature*, **361**, 88-91.

Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucl. Acids Res.*, **22**, 4673-4680.

Trott, O., Palmer, A.G. (2002). R1rho relaxation outside of the fast-exchange limit. *J. Magn. Res.*, **154**, 157-160.

van de Ven, F.J.M. (1995). *The basics of multi-dimensional NMR in liquids*. New York: Wiley-VCH.

Vermeulen, M., Carozza, M.J, Lasonder, E., Workman, J.L., Logie, C. & Stunnenberg, H.G. (2004). In vitro targeting reveals intrinsic histone tail specificity of the Sin3/histone deacetylase and N-CoR/SMRT corepressor complexes. *Mol. Cell. Biol.*, **24**, 2364-2372.

Viles, J.H, Donne, D., Kroon, G., Prusiner, S.B., Cohen, F.E., Dyson, H.J. & Wright, P.E. (2001). Local structural plasticity of the prion protein. Analysis of NMR relaxation dynamics. *Biochem.*, **40**, 2743-2753.

Volkman, B.F., Lipson, D., Wemmer, D.E. and Kern, D. (2001). Two-state allosteric behavior in a single-domain signaling protein. *Science*, **291**, 2429-2433.

Vriend, G. (1990). WHAT-IF: a molecular modelling and drug design program. *J. Mol. Graphics*, **8**, 52-56.

Vuister, G.W., Boelens, R., Padilla, A., Kleywegt, G.J. and Kaptein, R. (1990). Assignment strategies in homonuclear three-dimensional ^{1}H NMR spectra of proteins. *Biochem.*, **29**, 1829-1839.

Wang, H., Clark, I., Nicholson, P.R., Herskowitz, I. & Stillman, D.J. (1990). The Sacchoromyces cerevisae SIN3 gene, a negative regulator of HO, contains four amphipathic helix motifs. *Moll. Cell. Biol.*, **10**, 5927-2936.

Wijmenga, S.S., Mierlo, C.P.M. van and Steensma, E. (1996). Doubly sensitivity-enhanced 3D TOCSY-HSQC. *J. Biomol. NMR*, **8**, 319-330.

Wishart, D.S. and Sykes, B.D. (1994). The ^{13}C chemical shift method. A simple method for the identification of protein secondary structure using ^{13}C chemical shift data. *J. Biomol. NMR*, **4**, 171-180.

Woessner, D.E. (1961). Nuclear transfer effects in nuclear magnetic resonance pulse experiments, *J. Chem. Phys.*, **35**, 41-48.

Wüthrich, K. (1986). *NMR of proteins and nucleic acids*. New York: Wiley.

Xin, H., Yoon, H.G., Singh, P.B., Wong, J. and Qin, J. (2004). Components of a pathway maintaining histone modification and heterochromatin protein 1 binding at the pericentric heterochromatin in Mammalian cells. *J. Biol. Chem.*, **279**, 9539-9546.

Yang, L., Mei, Q., Zielinska-Kwiatkowska, A., Matsui, Y., Blackburn, M.L., Benedetti, D., Krumm, A.A., Taborsky, G.J. Jr. and Chansky, H.A. (2003). An ERG (ets-related gene)-associated histone methyltransferase interacts with histone deacetylases 1/2 and transcription co-repressors mSin3A/B, *Biochem. J.*, **369**, 651-657.

Yang, X.-J. and Seto, E. (2003). Collaborative spirit of histone deacetylases in regulating chromatin structure and gene expression. *Curr. Opin. Genet. Dev.*, **13**, 143-153.

Yang, X., Zhang, F. & Kudlow, J.E. (2002). Recruitment of O-GlcNAc transferase to promoters by corepressor mSin3A: coupling protein O-GlcNAcetylation to transcriptional repression. *Cell*, **110**, 69-80.

Yee, A.A., Savchenko, A., Ignachenko, A., Lukin, J., Xu, X., Skarina, T., Evdokimova, E., Liu, C.S., Semesi, A., Guido, V., Edwards, A.M. and Arrowsmith, C.H. (2005). NMR and X-ray crystallography, complementary tools in structural proteomics of small proteins. *J. Am. Chem. Soc.*, **127**, 16512-16517.

Zhang, M.J. and Yuan, T. (1998). Molecular mechanisms of calmodulin's functional versatility. Biochem. *Cell. Biol.*, **76**, 313-313.

Zhang, S. and Gorenstein, D.G. (2000). Synchronized adiabatic decoupling. *J. Magn. Reson.*, **147**, 110-115.

Zhang, Y., Iratni, R., Erdjument-Bromage, H., Tempst, P. & Reinberg, D. (1997). Histone deacetylases and SAP18, a novel polypeptide, are components of a human Sin3 complex. *Cell*, **89**, 357-364.

Zhang, Y., Sun, Z.W., Iratni, R., Erdjument-Bromage, H., Tempst, P., Hampsey, M. & Reinberg, D. (1998). SAP30, a novel protein conserved between human and yeast, is a component of a histone deacetylase complex. *Mol. Cell*, **1**, 1021-1031.

Zhou, P., Lugovskoy, A.A. & Wagner, G. (2001). A solubility enhancement tag (SET) for NMR studies of poorly behaving proteins. *J. Biomol. NMR*, **20**, 11-14.

Zhou, Z.Q. and Hurlin, P.J. (2001). The interplay between Mad and Myc in proliferation and differentiation. *Trends Cell Biol.*, **11**, S10-S14.

Zuiderweg, E.R. (2002). Mapping protein-protein interactions in solution by NMR spectroscopy. *Biochem.*, **41**, 1-7.

Summary
Samenvatting
Eenvoudige samenvatting
List of publications
Curriculum vitae
Dankwoord

Summary

To unravel life's molecular processes is the ultimate challenge for the life sciences. As nearly all these processes involve proteins, understanding of protein function is essential to meet this challenge. Protein function is intricately related to their molecular structure and dynamics. Therefore, the study of this structure-mobility-function relationship is one of the main goals in structural biology. Nuclear magnetic resonance (NMR) is especially well suited for such a study as it can be used to both determine the atomic, three-dimensional structure as well as the dynamics of proteins. This thesis concerns such a study on an important protein domain, the second PAH domain of a protein called Sin3, which is required for the development of an embryo to a healthy adult. The goal of this study was to describe how this domain interacts with other proteins and to develop new NMR methods for the study of such systems.

Chapter 1 gives an introduction of the Sin3 protein and its role in the repression of the transcription of many genes. Sin3 is recruited to its target genes via its second PAH domain, referred to as PAH2. The target gene is first recognized by a specific DNA-binding protein, which subsequently binds to the PAH2 domain of Sin3. The multitude of interactions involving the PAH2 domain make it an interesting subject to gain insight in protein-protein interactions. The chapter is concluded with a brief overview of the role of NMR in structural biology and of NMR theory.

Chapter 2 describes a specific type of experimental artifacts sometimes observed in multidimensional NMR spectra. It is shown that under specific conditions the heteronuclear decoupling scheme can generate ridges in the spectrum. These ridges occur at defined frequencies in the indirect dimensions, compromising the accurate analysis of peaks in the spectrum. A theoretical explanation of the artifact is presented and used to simulate an experimentally observed artifact in a two-dimensional spectrum. Finally, a simple way of preventing the artifact is described.

In the study of protein-peptide complexes, the peptide is usually studied using two-dimensional spectra. However, when the peptide is too large, these spectra can become too crowded. To overcome this problem, a three-dimensional TOCSY-TOCSY experiment for the study of unlabeled substrates in complex with a labeled protein is presented in **Chapter 3**. Using this new experiment, the complex of an extended peptide derived from the Sin3 Interacting Domain (SID) of the protein Mad1 and the PAH2 domain was successfully assigned.

The extension of the SID results in a higher binding affinity to the PAH2 domain. The three-dimensional structure of this complex, presented in **Chapter 4**, shows that two residues in the extension mediate additional interactions with the PAH2 domain. This demonstrates that residues outside the hydrophobic core of the SID can interact with the PAH domain and suggests that these additional interactions are important to modulate binding affinity and specificity to appropriate levels.

Understanding of the molecular mechanism of protein–protein interactions requires knowledge of both bound and unbound states of proteins. In **Chapter 5** the three-dimensional structure of the unbound PAH2 domain is presented. Both the general structure and the conformation of binding pocket are highly similar to that in the PAH2-SID complex. Combined with the unstructured nature of the unbound SID, this suggests that the PAH2 domain acts a folding template for the SID. The PAH1 and PAH3, but not the PAH4 domain, most likely have iden-

tical folds and modes of interaction, given their sequence similarity. Furthermore, this chapter describes that the unbound PAH2 domain is in equilibrium between the folded state and a partially unfolded minor state.

The structure of this minor state was characterized using a new NMR method, called CEESY, described in **Chapter 6**. Although its low population and transient nature preclude a direct observation of the minor state, the CEESY experiment enables its indirect structural characterization. Qualitative chemical shift information is obtained that can be translated into structural information by using statistical methods. As the CEESY experiment can be applied to multiple nuclei, it offers a detailed structural characterization of unobservable protein states.

Chapter 7 addresses the specificity of the PAH-SID interaction on the basis of the results described in this thesis and in the literature. Subtle sequence differences between the different PAH domains are identified that influence their basal level of specificity. Comparison of different PAH-SID structures shows that different SIDs have a different orientation and position to optimize their fit in the binding pocket. Furthermore, each SID is involved in unique additional interactions which further boost the specificity. Finally, a model describing complex formation is proposed and suggestions for further research are made.

Samenvatting

Het begrip van het leven op moleculair niveau is de ultieme uitdaging van de levenswetenschappen. Onderzoek naar eiwitten is hierin essentieel, omdat eiwitten in vrijwel alle cellulaire processen een belangrijke functie hebben. Deze functie is nauw verbonden met hun moleculaire structuur en dynamica. De studie van de structuur-mobiliteit-functie relatie is dan ook een hoofddoel in structurele biologie. Zulk onderzoek kan worden gedaan met behulp van kernspin magnetische resonantie (NMR), waarmee zowel de driedimensionale structuur als de dynamica van eiwitten bepaald kan worden. Dit proefschrift beschrijft een dergelijk onderzoek voor een belangrijk eiwit domein, het tweede PAH domein van het eiwit Sin3. Dit eiwit is noodzakelijk bij de ontwikkeling van een embryo tot een gezonde volwassene. Het doel van dit onderzoek was te beschrijven hoe dit domein met andere eiwitten interacties aangaat en om nieuwe NMR technieken voor de bestudering van dergelijke systemen te ontwikkelen.

In **Hoofdstuk 1** wordt het Sin3 eiwit geïntroduceerd en wordt beschreven hoe Sin3 de transcriptie van vele genen verhindert. Het tweede PAH, ofwel PAH2, domein van Sin3 speelt een belangrijke rol bij het vinden van deze genen op het DNA. Het gen wordt eerst herkend door een specifiek DNA-bindend eiwit, dat vervolgens aan het PAH2 domein van Sin3 bindt. Door de veelheid en verscheidenheid van interacties met verschillende DNA-bindende eiwitten, is het PAH2 domein een interessant onderwerp van onderzoek om ons inzicht in de principes van eiwit-eiwit interacties te vergroten. Het hoofdstuk besluit met een kort overzicht van de rol van NMR in structurele biologie en van NMR theorie.

Hoofdstuk 2 beschrijft een type artefact in multidimensionale NMR spectra. Het toont aan dat onder specifieke condities de heteronucleaire ontkoppelsequentie tot vervelende strepen in het spectrum kan leiden. Deze strepen komen voor op specifieke frequenties in de indirecte dimensie en kunnen daardoor de analyse van pieken in het spectrum verhinderen. Een theoretische verklaring voor het ontstaan van het artefact wordt gegeven en op basis hiervan wordt het experimenteel waargenomen artefact in een tweedimensionaal spectrum gesimuleerd. Tot slot wordt een simpele manier beschreven om het artefact te voorkomen.

In de studie van eiwit-peptide complexen wordt het peptide gewoonlijk met behulp van tweedimensionale spectra bestudeerd. Echter, wanneer het peptide te groot is, kan deze analyse vaak niet worden volbracht door overlap van de vele pieken in het spectrum. Om dit probleem op te lossen, wordt in **Hoofdstuk 3** een driedimensionaal TOCSY-TOCSY experiment beschreven voor de studie van ongelabelde substraten aan gelabelde eiwitten. Met behulp van dit experiment, is de resonantietoekenning van het complex van een verlengd peptide, corresponderend met het Sin3 bindende domein (SID) van het eiwit Mad1, en het PAH2 domein met succes afgerond.

Door de verlenging bindt deze SID sterker aan het PAH2 domein. De driedimensionale structuur van het complex, beschreven in **Hoofdstuk 4**, laat zien dat twee residuen in de verlenging betrokken zijn bij extra interacties met het PAH2 domein. Dit demonstreert dat residuen buiten de hydrofobe kern van de SID ook interacties kunnen hebben met het PAH2 domein en suggereert verder dat zulke additionele interacties een belangrijke rol kunnen spelen om de bindingsaffiniteit en specificiteit op het gewenste niveau te brengen.

Om eiwit-eiwit interacties te begrijpen, is kennis van zowel de ongebonden als de gebonden vorm van het eiwit nodig. In **Hoofdstuk 5** wordt de driedimensionale structuur van het ongebonden PAH2 domein beschreven. Zowel de algemene vorm als de precieze conformatie van de oppervlakte waaraan de SID bindt, zijn vrijwel gelijk aan dat in de gebonden vorm. Aangezien de ongebonden SID ongestructureerd is, suggereert dit dat het PAH2 domein dient als een mal voor de correcte structurering van de SID. Ook het PAH1 en PAH3, maar niet het PAH4 domein hebben waarschijnlijk een gelijke structuur en interactie modus, gegeven hun overeenkomst in sequentie. Verder wordt in dit hoofdstuk aangetoond dat het ongebonden PAH2 domein voor een fractie van de tijd in een gedeeltelijk ontvouwen neventoestand verkeert.

De structuur van deze neventoestand werd gekarakteriseerd met een nieuwe NMR methode, CEESY, beschreven in **Hoofdstuk 6**. Hoewel de lage populatie en korte levensduur van de neventoestand een direct observatie onmogelijk maken, kan met het CEESY experiment deze toestand toch worden gekarakteriseerd. CEESY geeft kwalitatieve informatie over de chemical shift die vervolgens omgezet kan worden in structurele informatie met behulp van statistische methoden. Omdat het CEESY experiment op verschillende atoomkernen kan worden toegepast, biedt het de mogelijkheid om tot een gedetailleerde karakterisatie van de structuur van ‘onzichtbare’ toestanden van eiwitten te komen.

Hoofdstuk 7 behandelt de specificiteit van de PAH-SID interactie op basis van de resultaten beschreven in dit proefschrift en in de literatuur. Subtiele verschillen in sequentie tussen de verschillende PAH domeinen beïnvloeden het basale niveau van specificiteit. Vergelijking van verschillende PAH-SID complexen laat zien dat de verschillende SIDs een verschillende oriëntatie en positie hebben om zich zo optimaal mogelijk aan te passen aan de precieze vorm van de bindingsoppervlakte. Daarnaast heeft iedere SID ook unieke, additionele interacties die de specificiteit verder bevorderen. Tot slot wordt een model voorgesteld om de vorming van het PAH-SID complex te beschrijven en worden suggesties voor verder onderzoek gedaan.

Eenvoudige samenvatting

Onderdrukking is goed, zelfs noodzakelijk. Als het tenminste om onze erfelijke informatie gaat. Onze erfelijke informatie bevat namelijk de blauwdruk van ons hele lichaam en is in iedere cel van ons lichaam aanwezig. Het is duidelijk dat één enkele cel niet de hele blauwdruk nodig heeft, maar slechts een klein stukje daarvan. En ieder type cel heeft weer een ander stukje nodig; een hersencel, een levercel en een spiercel moeten verschillende dingen doen en lezen dus verschillende gedeelten van de blauwdruk. Dit betekent dat delen van onze erfelijke informatie die voor een cel niet nodig zijn onderdrukt moeten worden. Wanneer deze onderdrukking niet goed werkt, kan dit nare gevolgen hebben, zoals het ontstaan van kanker en ontwikkelingsziekten.

In dit onderzoek is een klein gedeelte onderzocht van het mechanisme om de onnodige informatie te onderdrukken, namelijk het tweede PAH domein van het eiwit Sin3. Eiwitten zijn de 'werkpaarden' van de cel en voeren allerlei taken uit zoals het doorgeven van boodschappen, transport van voedingsstoffen en de regulering van allerlei processen, zoals het onderdrukken van onnodige erfelijke informatie. Ons erfelijk materiaal is georganiseerd in kleine stukjes, genen genoemd. Het eiwit Sin3 vormt de basis van een complex van eiwitten die gezamenlijk de onnodige genen onderdrukken. Sin3 laat zich vergelijken met een kapstok waaraan de andere eiwitten in het complex zich vastbinden. Het tweede PAH domein (PAH2) is een gedeelte van het Sin3 eiwit en vormt als het ware één van de haken van de kapstop. Hoe dit eiwit-complex de genen onderdrukt en de rol van het PAH2 domein hierin, wordt in Hoofdstuk 1 beschreven (zie ook box 1).

Box 0. Scheikunde: atomen & moleculen

Wij zijn geen zuivere koffie, zou je kunnen zeggen. Een mens is voor een scheikundige een complex mengsel van allerlei stoffen: vooral water, wat vet, een snufje zout, natuurlijk wat DNA en veel eiwitten. Op grond van vele observaties en experimenten kunnen we niet anders dan aannemen dat alle stoffen zijn opgebouwd uit atomen, waarvan er maar zo'n 110 verschillende typen bestaan. Ook is het logischerwijs noodzakelijk dat veel stoffen zijn opgebouwd uit moleculen, verbindingen van verschillende atomen. Bijvoorbeeld, op basis van allerlei experimenten moet water zijn opgebouwd uit moleculen die ieder uit één atoom zuurstof en twee atomen waterstof, vandaar de welbekende formule H_2O. Eiwitten bestaan uit veel ingewikkeldere moleculen die wel uit tienduizenden atomen kunnen bestaan: bv. de formule voor het Sin3 eiwit is $C_{5579}H_{8750}N_{1598}O_{1678}S_{42}$.

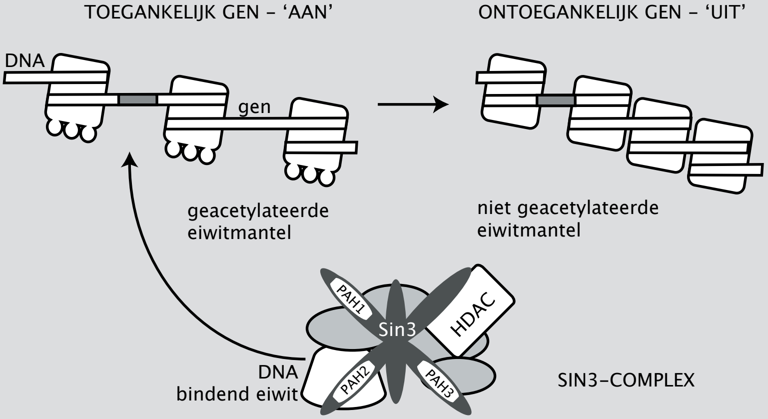
In dit proefschrift is de binding tussen het PAH2 domein van Sin3 en andere eiwitten onderzocht met behulp van kernspinresonantie (afgekort NMR, zie ook box 2). Met behulp van deze techniek kan zowel de driedimensionale vorm als de bewegelijkheid van eiwitten worden bepaald, wat leidt tot een beter begrip van de werking van eiwitten. Deze kennis kan vervolgens als basis dienen voor de ontwikkeling van nieuwe medicijnen of van nieuwe materialen.

Ik heb de vorm en bewegelijkheid van het PAH2 domein bepaald (Hoofdstuk 5, zie ook box 3) en van de binding tussen het PAH2 domein en een van de andere eiwitten uit het Sin3-complex (Hoofdstuk 4). Dit was alleen mogelijk door behalve bestaande NMR technieken te gebruiken, ook nieuwe experimenten te ontwikkelen. Zo werden storingen verwijderd in een experiment om de bewegelijkheid van eiwitten te meten (Hoofdstuk 2), werd het onderzoek naar de binding tussen eiwitten met een complexe samenstelling vergemakkelijkt (Hoofdstuk3) en werd het mogelijk om instabiele vormen van eiwitten te karakteriseren (Hoofdstuk 6).

Box 1. Het Sin3-complex & de onderdrukking van de genetische informatie

Onze genetische informatie is opgeslagen op het deoxyribonuleïnezuur molecuul, beter bekend als het DNA. Het DNA is opgebouwd uit vier bouwstenen. De volgorde van deze bouwstenen is de genetische informatie. In een gen bepaalt de volgorde van deze bouwstenen welk eiwit wordt aangemaakt. In de eerste stap van de eiwit-productie, transcriptie, wordt deze volgorde afgelezen. Vervolgens wordt deze vertaald in de aminozuurvolgorde van het eiwit.

Het Sin3-complex onderdrukt de transcriptie van genen door te zorgen dat het gen ontoegankelijk wordt en dus niet meer kan worden afgelezen (zie Figuur 1). Het ontoegankelijk maken gebeurt door de eiwitmantel van het DNA te manipuleren. Het Sin3-complex onderdrukt genen waarvan de eiwitten betrokken zijn bij processen zoals de regulering van de celcyclus en van de embryonale ontwikkeling. Ieder van deze genen wordt herkend door een gespecialiseerd DNA bindend eiwit. Dit eiwit bindt aan een PAH domein van Sin3. Zodoende wordt het gehele Sin3-complex bij het gen gebracht. Door een ander gespecialiseerd eiwit, het zogenaamde HDAC eiwit, wordt vervolgens de eiwitmantel gemanipuleerd zodat het gen ontoegankelijk wordt en dus niet meer kan worden afgelezen.



Figuur 1. Schema van de onderdrukking van de transcriptie door het Sin3 eiwit. De PAH domeinen van Sin3 zijn aangegeven. Sin3 is onder andere gebonden aan een DNA-bindend eiwit, dat een specifieke DNA sequentie (grijze stuk op het gen) herkent, en een histon deacetylase enzym (HDAC). Menselijk DNA is omgeven door een eiwitmantel, ook wel histonen genoemd. Als deze mantel geacetyleerd is, is het gen toegankelijk en kan transcriptie plaatsvinden. Echter, wanneer deze acetyl-groep verwijderd wordt door de HDAC enzym, wordt het gen ontoegankelijk en kan er geen transcriptie meer plaatsvinden.

Mijn resultaten geven inzicht in hoe het PAH2 domein aan andere eiwitten bindt. Zo is het nu tegen de verwachting in duidelijk dat het PAH2 domein niet van vorm verandert als het aan een ander eiwit bindt. Hierdoor is het waarschijnlijk dat het PAH2 domein als een soort mal dient, en de vorm van het partner-eiwit bepaalt. Het PAH2 domein klemt het partner-eiwit stevig vast wanneer ze aan elkaar binden. Nu is vastgesteld dat het contactoppervlak tussen beide aanzienlijk groter is dan tot nog toe was aangenomen. Op basis van deze observaties en die van collega-wetenschappers is een model geformuleerd om de werking van de PAH domeinen te beschrijven (Hoofdstuk 7).

Deze resultaten dragen niet alleen bij aan ons begrip van de werking van het Sin3 eiwit, het draagt ook bij aan ons begrip van interacties tussen eiwitten. En nu snappen we beter hoe erfelijke informatie, voor ons eigen bestwil, wordt onderdrukt.

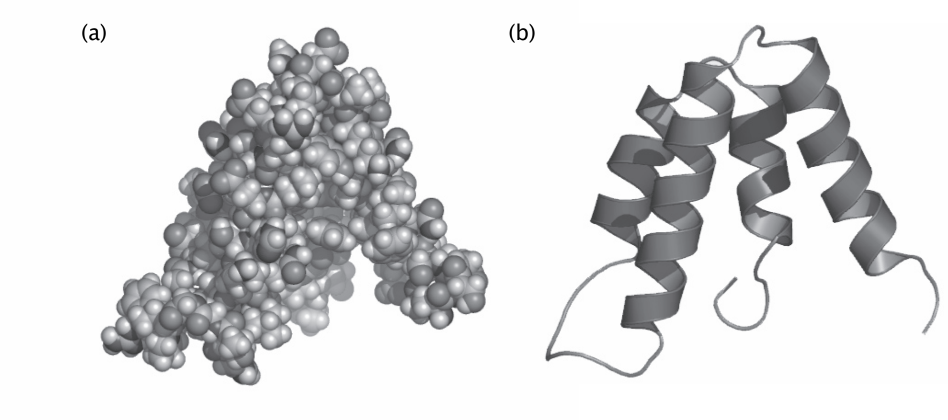
Box 2. Kernspinresonantie (NMR): ‘atomen interviewen’

Kernspinresonantie (in het Engels nuclear magnetic resonance, afgekort NMR) is een prachtige techniek om de structuur van moleculen te onderzoeken. Het mooie van NMR is dat je als onderzoeker een bepaalde mate van invloed kunt uitoefenen op het molecuul om specifieke informatie te verkrijgen. Op een geheel onschadelijke wijze kun je als het ware de atomen ‘interviewen’ over hun directe omgeving; je kunt zo informatie vergaren over het type atomen in de buurt van het geïnterviewde atoom of over de mate van bewegelijkheid op die plek in het molecuul. Door heel veel atomen in het molecuul te ‘interviewen’ en alle informatie te analyseren, kun je uiteindelijk tot een goed beeld van de vorm en bewegelijkheid van het hele molecuul komen, zelfs voor grote moleculen zoals eiwitten.

NMR is een vorm van spectroscopie (letterlijk: ‘kijken naar licht’), wat betekent dat de eigenschappen van moleculen worden onderzocht met behulp van een of andere vorm van straling. In het geval van NMR wordt dezelfde straling gebruikt als voor de radio en wordt gebruik gemaakt van de magnetische eigenschappen van sommige atomen, de zogenaamde kernspin. In een NMR experiment wordt de te onderzoeken stof in een sterk magneetveld gebracht om te zorgen voor voldoende volume en helderheid in het antwoord van de kernspins. Vervolgens wordt volgens een strak gedefinieerd ritme een aantal zeer korte pulsen van radiogolven op de stof afgevuurd, de vraag van de onderzoeker. Deze vraag ‘resoneert’ bij de kernspins en leidt tot een respons. Dit antwoord wordt opgenomen en geanalyseerd. Door slimme vragen te stellen kun je ondanks de beperkte woordenschat van de kernspins toch heel veel gedetailleerde informatie verkrijgen.

Box 3. Eiwit in 3D

Alle eiwitten zijn opgebouwd uit 20 verschillende aminozuren. Eiwitten worden gemaakt door deze aminozuren in een specifieke volgorde aan elkaar te koppelen, zoals in het gen is gedefinieerd. De 20 verschillende aminozuren bestaan allemaal uit eenzelfde basisgedeelte maar hebben ieder ook een variabele groep, de zijketen genaamd. Door de aminozuren aan elkaar te koppelen ontstaat een lange hoofdketen met daaraan de verschillende zijketens. Het eiwit neemt daarna zijn unieke vorm aan die bijvoorbeeld met behulp van NMR bepaald kan worden (zie Figuur 2).



Figuur 2. Verschillende voorstellingen van de vorm van het PAH2 domein (a). voorstelling waarbij alle atomen worden voorgesteld als bolletjes. (b). alleen de vorm van de hoofdketen wordt schematisch weergegeven. Hierin is de typische schroefstructuur van delen van de hoofdketen goed te zien.

List of publications

1. van Ingen, H., Vuister, G.W. and Tessari, M. (2002). A two-dimensional artifact from asynchronous decoupling. *J. Magn. Reson.*, **156**, 258-261.

2. van Ingen, H., Tessari, M. and Vuister, G.W. (2002). A 3D doubly sensitivity enhanced X-filtered TOCSY-TOCSY experiment. *J. Biomol. NMR*, **24**, 155-160.

3. Wingens, M., Walma, T., van Ingen, H., Stortelers, C., van Leeuwen, J.E., van Zoelen, E.J. and Vuister, G.W. (2003). Structural analysis of an epidermal growth factor/transforming growth factor-alpha chimera with unique ErbB binding specificity. *J. Biol. Chem.*, **278**, 39114-39123.

4. van Ingen, H., Lasonder, E., Jansen, J.F.A., Kaan, A.M., Spronk, C.A.E.M., Stunnenberg, H.G.M. & Vuister, G.W. (2004). Extension of the binding motif of the Sin3 interacting domain of the Mad family proteins. *Biochem.*, **43**, 46-54.

5. van Ingen, H., Baltussen, M.A.H., Aelen, J.M.A, Vuister, G.W. (2006). Role of structural and dynamical plasticity in Sin3: the free PAH2 domain is a folded module in mSin3B. *J. Mol. Biol.*, **358**, 485-497.

6. van Ingen, H., Vuister, G.W., Wijmega, S. & Tessari, M. (2006). CEESY: Characterizing the conformation of unobservable protein states. *J. Am. Chem. Soc.*, **128**, 3856-3858.

7. van Ingen, H. and Vuister, G.W. Specificity in the PAH-SID interaction. *manuscript in preparation*.

Curriculum vitae

Hugo van Ingen is op 30 november 1976 in Weurt geboren als zoon van Cor en Leny en broer van Linda van Ingen. In 1994 behaalde hij zijn VWO Gymnasium diploma aan het Dominicus College in Nijmegen. In dat jaar begon hij aan zijn studie Scheikunde aan de Radboud Universiteit Nijmegen, destijds de Katholieke Universiteit Nijmegen. Tijdens deze studie heeft hij zijn hoofdvakstage gedaan bij de afdeling Fysische Chemie/Biofysische Chemie onder leiding van dr. T. Papavoine en prof. dr. C.W. Hilbers. Hij ontving voor dit onderzoek de Unilever Research prijs. Zijn bijvakstage deed hij bij Psychologie alwaar hij zijn liefde Maja Stulemeijer leerde kennen. Het doctoraal examen Scheikunde werd cum laude behaald in december 1999.

In 2000 volgde Hugo de postdoctorale lerarenopleiding Scheikunde aan het Instituut voor Leraar en School in Nijmegen, waar hij zijn eerstegraads lesbevoegdheid Scheikunde behaalde.

Van januari 2001 tot januari 2006 was hij werkzaam als junior onderzoeker bij de afdeling Fysische Chemie/Biofysische Chemie van de Radboud Universiteit onder begeleiding van dr. G.W. Vuister. In die periode werkte hij voor vier dagen in de week aan het onderzoek, beschreven in dit proefschrift. Tot april 2002 werkte hij voor één dag in de week als docent op de middelbare school. Daarna werkte hij voor één dag in de week als medewerker onderwijsontwikkeling voor het onderwijscluster Moleculaire Wetenschappen van de Faculteit Natuurwetenschappen, Wiskunde en Informatica

Sinds 1 maart 2006 is Hugo werkzaam als postdoctoraal onderzoeker bij de vakgroep NMR Spectroscopie van Universiteit Utrecht.

Dankwoord

Het is tijd voor het laatste loodje. Niet zonder reden ook het meest gelezen gedeelte van een proefschrift: ook al is het doen van onderzoek veel solo-werk, veel mensen hebben direct of indirect bijgedragen aan het tot stand komen van dit proefschrift.

Allereerst, Geerten, bedankt voor je begeleiding, niet alleen van het pure onderzoek, maar ook van mijn ontwikkeling als onderzoeker. Bedankt voor de vrijheid en de uitdagingen die je me hebt gegeven. Ik heb veel van je geleerd en ik bewonder je snelheid, scherpte en veerkracht. Ik wens je heel veel succes. We hadden ook nog een paar weddenschappen staan, volgens mij heb ik daar één van gewonnen, daar moeten we het nog maar eens over hebben...

Prof. Hilbers, ik blij dat u mijn promotor wilde zijn en vereerd dat ik nu uw laatste promovendus ben. Bedankt voor de stimulerende discussies, vooral in deze laatste fase.

Dit project is begonnen als een samenwerking met de afdeling Moleculaire Biologie van prof. Stunnenberg. Ik wil mijn co-auteurs bedanken voor hun waardevolle input en inzet. A special thanks to Xavier, for your expertise in the biology of Sin3/PAH.

Marco, zonder jou was dit boekje dunner geweest. Ik heb onze samenwerking heel erg gewaardeerd. Ik bewonder je kennis van zaken en je enorme toewijding. Ik heb veel van je geleerd en ik hoop dat je ook een beetje van mij hebt opgestoken.

Jan A., zonder eiwit geen onderzoek. Bedankt voor jouw inzet, je hulp en de prettige samenwerking. Frank en Aafke, bedankt voor jullie hulp bij mijn sporadische labactiviteiten.

Jos, bedankt voor de zorg voor de computers en de spectrometers. Ook de oscilloscoop en expertise van Jan v. O. en Hans was soms onmisbaar. Harrie, jouw vrolijkheid en daadkracht maakten een verschil. Babs, bedankt voor je goede zorgen en samen met Marian en Maaike bedankt voor de secretariële ondersteuning. Sybren, bedankt voor je interesse en de prettige samenwerking. Arno, bedankt voor de mogelijkheid om binnen de faculteit mijn onderwijsambities levendig te kunnen houden.

Mijn studenten, Maria en Marlies, bedankt voor jullie inzet tijdens de stage. Het was een plezierige tijd met veel welkome extra interactie. Maria, heel knap hoe je je stage hebt gecombineerd met een druk gezinsleven en de geboorte van Mats. Marlies, heel leuk dat je verder bent gegaan met magnetische resonantie en heel veel succes met jouw promotie!

Ik ben zo lang op de afdeling geweest dat ik verschillende generaties 'vloeistoffers' en 'vaste stoffers' heb te bedanken voor hulp, discussies, een prettige werksfeer, gezelligheid en de onvermijdelijke verjaardagstaart!

Paul, Chris en Cathelijne, jullie hebben als veteranen van de afdeling de juiste toon gezet. Samen met Tine ben ik verhuisd vanuit de donkere kelders van het voormalige UL naar een heuse kamer met uitzicht. Tine, thanks for your help and the great 'potluck' dinners. Miriam, je hebt ons daar een hele tijd gezelschap gehouden en dankzij jou heb ik de botanische tuin ontdekt. Met de komst van Sybren kwamen ook Fred, Bin en Chintaka. Guys, you were an important new impulse to the department. I really enjoyed our time together. Fred en Andreas, bedankt voor jullie vriendschap, de vele gesprekken over de afdeling en de Refter-maaltijden als er weer eens doorgewerkt moest worden. Ik ben blij dat jullie paranimf willen zijn. Fred, veel succes met Spinnovation! Andreas, door jou is mijn muzieksmaak weer verder verbreed. Mark, bedankt voor al je hulp, de vele discussies en de prettige tijd. En dankzij jou werk ik iedere dag weer met veel plezier met het meest prettige OS, ! Also thanks to all Marie-Curie students who brought an international and lively flavor to Nijmegen. Ramon en Otmar, de nieuwste lichting, bedankt voor de plezierige tijd in de kantoortuin en veel succes met jullie boekje! En houd die blauwe pixels in de gaten!

Alle vrienden, het clubje oud-studiegenoten en de schaatsers, bedankt voor jullie interesse en de broodnodige ontspanning die ik bij jullie vond tijdens etentjes, spelletjes, weekenden en sportieve inspanningen op de fiets, schaats of skate.
Sinds de laatste jaren is de woensdagavond voor plan B. Onze CD dreigt bijna een soort proefschrift te worden, maar ook bij dit proefschrift heb ik het ervaren: je denkt het komt nooit af en dan, opeens, is het klaar...

Maarten en Vera, Jiri en Cynthia, bedankt voor jullie interesse en medeleven. Na mijn promotie nog ééntje en dan hebben jullie vier doctoren in de familie!
Papa en mama, het is een cliché, maar dat komt omdat het waar is: zonder jullie was dit boekje er nooit geweest. Jullie hebben me altijd gestimuleerd en de ruimte gegeven om mijn fascinaties te volgen. Bedankt voor jullie steun en interesse in deze promotietijd.
Linda, ook jij bedankt voor je belangstelling en je vragen over wat ik nou eigenlijk doe. Ik hoop dat ik het nu eindelijk een beetje heb kunnen uitleggen.
Maja, lief, bedankt voor alles. Bedankt voor je geduld als ik weer eens door wilde werken, bedankt voor je steun als ik het weer eens niet zag zitten, bedankt voor de balans die je in mijn leven geeft. Straks als jij je boekje aan het schrijven bent, dan kun je op mij rekenen.